

QA 298. C67  
2005

RESERVADO

**UNIVERSIDADE TÉCNICA DE LISBOA**  
**INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO**

**Mestrado em: Econometria Aplicada e Previsão**

**APROXIMAÇÕES ASSIMPTÓTICAS NA ANÁLISE DA  
HETEROGENEIDADE NEGLIGENCIADA EM MODELOS  
DE DURAÇÃO**

**PAULO RENATO FREITAS COSTA**

**Orientação:** Prof. Doutor Montezuma Boaventura Guimarães Dumangane

**Juri**

**Presidente:** Prof. Doutor José Manuel de Matos Passos

**Vogais:** Prof. Doutora Esmeralda Jesus Ratinho Lopes Arranhado Ramalho  
Prof. Doutor Montezuma Boaventura Guimarães Dumangane

Março / 2005

# Resumo

Na análise econométrica de modelos de duração, é bastante comum depararmos com a existência de heterogeneidade negligenciada. Este problema deve-se, em grande parte, ao facto de não ser possível observar todas as características individuais que afectam a duração de um evento.

Neste trabalho, utilizando aproximações assintóticas, vamos estudar o impacto da heterogeneidade em modelos de duração com parâmetro de escala. Como as aproximações são sensíveis ao efeito da heterogeneidade, irão constituir a base para a construção de um teste para a sua detecção. Uma expressão que caracterize o enviesamento causado pela heterogeneidade será derivada utilizando as aproximações, sendo essa expressão a base para a construção de um estimador GMM que corrige o enviesamento na condição de momentos do modelo que ignora a heterogeneidade.

Utilizando duas distribuições paramétricas, Weibull e Log-logistic, e através de simulações de Monte Carlo, iremos analisar o comportamento do estimador GMM proposto.

**Palavras chave:** heterogeneidade negligenciada, modelos de duração, aproximações assintóticas, estimador GMM, simulação de Monte Carlo, teste score.

**Classificação JEL:** C12, C14, C15, C41

# Abstract

In the econometric analysis of duration models, it is common to come across itself with the problem of neglected heterogeneity. This problem, to a large extent, is due to the fact of not being possible to observe all the individual characteristics that affect the duration of an event.

In this work, small parameters approximations will be used to study the impact of the heterogeneity in duration models with a scale parameter. The approximations are sensible to the effect of the heterogeneity so, they will be the basis to construct a test for its detection. An expression that characterizes the bias caused by neglecting heterogeneity will be derived using the approximations, being this expression the base for the construction of a GMM that corrects the bias in the moment conditions of the model that ignores heterogeneity.

Using two parametric distributions, Weibull and Log-logistic, and through Monte Carlo simulation, we will analyze the performance of the proposed GMM estimator.

**Keywords:** neglected heterogeneity, duration models, asymptotic approximations, GMM estimator, Monte Carlo simulation, score test.

**JEL Classification:** C12, C14, C15, C41

# Índice

Lista de tabelas .....	6
Lista de figuras .....	7
Agradecimentos .....	8
<b>1 Introdução .....</b>	<b>9</b>
<b>2 Conceitos básicos.....</b>	<b>14</b>
2.1 Modelos de duração .....	14
2.2 Censura .....	16
<b>3 Heterogeneidade negligenciada em modelos de duração .....</b>	<b>18</b>
3.1 Modelo mistura <i>hazard</i> proporcional .....	19
3.1.1 Análise utilizando a função <i>hazard</i> marginal .....	20
3.1.2 Análise utilizando aproximações assintóticas .....	22
3.2 Modelo com heterogeneidade no parâmetro de escala .....	25
3.3 Heterogeneidade e erro de medida.....	27
<b>4 Modelos com uma duração .....</b>	<b>29</b>
4.1 Teste à heterogeneidade negligenciada .....	29
4.2 Enviesamento dos estimadores .....	32
4.2.1 Observações completas .....	32
4.2.2 Observações censuradas .....	36
4.2.3 Exemplos .....	38
4.3 Estimador GMM corrigido .....	41
<b>5 Modelos com múltiplas durações.....</b>	<b>47</b>
5.1 Aproximações lineares .....	47
5.2 Teste à heterogeneidade negligenciada .....	49
5.3 Enviesamento dos estimadores .....	51
5.3.1 Exemplos .....	53
5.4 Estimador GMM corrigido .....	57

<b>6 Simulações .....</b>	<b>58</b>
6.1 Modelo com observações completas sem regressores .....	61
6.2 Modelo com observações completas com um regressor.....	68
6.3 Modelo com observações censuradas .....	76
<b>7 Conclusões .....</b>	<b>83</b>
<b>A Aproximação da função condicional de enviesamento - uma     duração .....</b>	<b>85</b>
<b>B Aproximação da função condicional de enviesamento - múltiplas     durações .....</b>	<b>89</b>
<b>Bibliografia .....</b>	<b>91</b>

# Lista de tabelas

6.1	Parâmetros das distribuições de heterogeneidade .....	59
6.2	Resultados da simulação dos modelos Weibull e Log-logistic sem regressores: estimador de Máxima verosimilhança .....	62
6.3	Resultados da simulação dos modelos Weibull e Log-logistic sem regressores: estimador GMMR .....	63
6.4	Resultados da simulação do modelo Weibull sem regressores: estimador GMMU .....	64
6.5	Resultados da simulação do modelo Log-logistic sem regressores: estimador GMMU .....	65
6.6	Resultados da simulação dos modelos Weibull e Log-logistic com regressores para $\rho=0.8$ e $\alpha=1$ : estimador de Máxima Verosimilhança .....	69
6.7	Resultados da simulação dos modelos Weibull e Log-logistic com regressores para $\rho=0.8$ e $\alpha=1$ : estimador GMMR .....	70
6.8	Resultados da simulação do modelo Weibull com regressores para $\rho=0.8$ e $\alpha=1$ : estimador GMMU .....	71
6.9	Resultados da simulação do modelo Log-logistic com regressores para $\rho=0.8$ e $\alpha=1$ : estimador GMMU .....	72
6.10	Resultados da simulação dos modelos Weibull e Log-logistic com censura para $\rho=0.8$ e $\alpha=1$ : estimador de Máxima Verosimilhança .....	78
6.11	Resultados da simulação dos modelos Weibull e Log-logistic com censura para $\rho=0.8$ e $\alpha=1$ : estimador GMMR .....	79
6.12	Resultados da simulação o modelo Weibul com censura para $\rho=0.8$ e $\alpha=1$ : estimador GMMU .....	79
6.13	Resultados da simulação do modelo Log-logistic com censura para $\rho=0.8$ e $\alpha=1$ : estimador GMMU .....	80

# Lista de figuras

4.1	Enviesamento das condições de momentos sem correcção (pontos) e com correcção (tracejado) para a distribuição Weibull com distribuição da heterogeneidade Gamma (painéis da esquerda) e Lognormal (painéis da direita) .....	44
4.2	Enviesamento das condições de momentos sem correcção (pontos) e com correcção (tracejado) para a distribuição Log-logistic com distribuição da heterogeneidade Gamma (painéis da esquerda) e Lognormal (painéis da direita) .....	45
6.1	Densidades dos estimadores MLE, GMMR e GMMU da estimação de $\alpha$ no modelo Weibull com distribuição de heterogeneidade Lognormal e $n = 500$ .....	66
6.2	Densidades dos estimadores MLE, GMMR e GMMU da estimação de $\alpha$ no modelo Log-logistic com distribuição de heterogeneidade Lognormal e $n = 500$ .....	67
6.3	Densidades dos estimadores MLE, GMMR e GMMU da estimação de $\beta_1$ no modelo Weibull com distribuição de heterogeneidade Lognormal e $n = 500$ .....	74
6.4	Densidades dos estimadores MLE, GMMR e GMMU da estimação de $\beta_1$ no modelo Log-logistic com distribuição de heterogeneidade Lognormal e $n = 500$ .....	75
6.5	Densidades dos estimadores MLE, GMMR e GMMU da estimação de $\alpha$ e $\beta_1$ no modelo Weibull com censura e com distribuição de heterogeneidade Lognormal e $n = 500$ .....	81
6.6	Densidades dos estimadores MLE, GMMR e GMMU da estimação de $\alpha$ e $\beta_1$ no modelo Log-logistic com censura e com distribuição de heterogeneidade Lognormal e $n = 500$ .....	82

# Agradecimentos

Depois deste longo período de trabalho não posso deixar de agradecer a todas as pessoas que, directa ou indirectamente, me apoiaram na elaboração deste estudo.

Gostaria de expressar a minha gratidão ao meu orientador, Prof. Doutor Montezuma Boaventura Guimarães Dumangane, por toda a atenção dispensada e rigor exigido na realização desta tese.

Agradeço também à minha namorada, familiares e amigos por todo o apoio e incentivo que me deram na realização deste trabalho.



# 1. Introdução

A utilização de modelos de duração em estudos económicos tem sido muito frequente nos últimos anos. São exemplos de estudos a duração do desemprego (Lancaster, 1979, Devine e Kiefer, 1991 e Narendranathan, Nickell e Stern, 1985), a duração do casamento (Lillard, 1993), a duração de greves (Kennan, 1985), a duração da emigração (Lindstrom, 1996).

Estes modelos analisam o tempo que um indivíduo permanece num determinado estado, como por exemplo, o tempo que um indivíduo esteve desempregado. O estudo destes modelos é baseado na análise da função *hazard*, função que representa, num determinado momento  $t$ , a taxa de saída de um indivíduo do estado em que se encontra, sabendo que ele se manteve nesse estado até esse momento  $t$ .

Como os indivíduos em análise nos modelos de duração não são homogêneos, torna-se necessário incluir no modelo regressores que representem as características individuais que têm influência sobre o tempo de permanência dos indivíduos no estado. Assim, será a função *hazard* condicional nesses regressores a base da nossa análise.

Em modelos cuja distribuição da duração tem parâmetro de escala, o conjunto dos regressores individuais é incluído como parte integrante desse parâmetro de escala, tendo como impacto na função *hazard* acelerar ou desacelerar a taxa de saída dos indivíduos do estado em que se encontram.

No entanto, é impossível a inclusão nos regressores de todas as variáveis que possam influenciar a duração, principalmente porque existem muitas características individuais que não são observáveis. É o caso da habilidade para uma determinada actividade ou para a procura de emprego, que são características que vão influen-

ciar a taxa de saída do desemprego de um indivíduo mas que não são observáveis. Ao considerarmos a heterogeneidade negligenciada como o conjunto dessas variáveis, estamos a colocar a heterogeneidade como um problema de omissão de regressores.

Lancaster (1979) foi o primeiro a abordar a questão da omissão de regressores, tendo demonstrado que essa omissão origina o enviesamento dos parâmetros funcionais da distribuição da duração em direcção a zero. Como consequência desse enviesamento, em média, os indivíduos com "elevada" heterogeneidade vão ter uma taxa de saída do estado superior aos indivíduos com "pequena" heterogeneidade.

Para evitar esse enviesamento é necessário incluir o efeito da heterogeneidade em cada indivíduo. Para tal, considera-se que a heterogeneidade é a realização de uma variável aleatória com uma determinada distribuição de probabilidade, sendo independente do tempo e dos regressores, assim como entre diferentes indivíduos. A sua inclusão no modelo é feita assumindo que essa variável é um regressor.

Na análise de modelos de duração, são muito utilizados os modelos da família das funções *hazard* proporcionais (ver Cox, 1972). Nestes modelos, a função *hazard* condicional nos regressores é composta por duas componentes que entram multiplicativamente na função, sendo dada por

$$\lambda(t, x) = \lambda_0(t) \varphi(x)$$

onde  $\lambda_0(t)$ , a *baseline hazard*, representa a forma funcional da função *hazard* e  $\varphi(x)$  o conjunto dos regressores, sendo usual especificar  $\varphi(x) = \exp(x'\beta)$ . Nestes modelos, incluir a variável da heterogeneidade como um regressor é equivalente a incluir essa variável de forma multiplicativa com a função *hazard*, dando origem ao modelo mistura. Tem sido este o modelo preferido para a estimação de modelos de duração com heterogeneidade negligenciada (ver Van den Berg, 2001).

Heckman e Singer (1984) e Lancaster (1990) demonstraram que, em modelos mistura, a elasticidade da função *hazard* onde a heterogeneidade é ignorada é superior à elasticidade da função *hazard* que inclui a heterogeneidade. Lancaster (1985) mostrou que na estimação do modelo Weibull onde se ignora a heterogeneidade, a estimação dos seus parâmetros é enviesada em direcção a zero.

No entanto, coloca-se o problema de qual a distribuição paramétrica da heterogeneidade que devemos escolher. Se por um lado, não há argumentos que nos levem a assumir uma distribuição em detrimento de outra, tendo mesmo trabalhos realizados por Heckman e Taber (1994) e Horowitz (1999) indicado que os dados de duração fornecem pouca informação sobre as características dessa distribuição, por outro lado, Heckman e Singer (1984) demonstraram que os estimadores do modelo mistura são sensíveis à forma funcional da distribuição da heterogeneidade.

Horowitz (1999) estimou semiparametricamente o modelo mistura sem especificar a distribuição da heterogeneidade. No entanto, na estimação paramétrica e semiparamétrica em que é definida uma distribuição para a heterogeneidade, coloca-se o problema de qual a distribuição da heterogeneidade que se deve assumir. Lancaster (1979) assumiu uma distribuição Gamma, Hougaard (1984) uma distribuição Normal inversa, Nikell (1979) uma distribuição Discreta, tendo outros autores tomado outras decisões (ver Van den Berg, 2001).

Apesar de existirem opiniões diversas sobre qual a distribuição da heterogeneidade que se deve assumir, tem sido a distribuição Gamma a que tem ganho mais adeptos na estimação do modelo mistura proporcional. Esta preferência deve-se, em grande parte, ao facto da função marginal obtida, após integrar a distribuição da duração com uma função densidade Gamma, ter uma solução fechada. Mais recentemente, Abbring e Van den Berg (2003) mostraram que a distribuição da heterogenei-

dade converge para uma distribuição Gamma.

Porém, em modelos de duração com parâmetro de escala não pertencentes à família das funções *hazard* proporcionais, o efeito da heterogeneidade pode não ser o mesmo que o verificado anteriormente, podendo mesmo a heterogeneidade ter efeitos diferentes na função *hazard* ao longo do tempo.

Neste trabalho, utilizando aproximações assintóticas, vamos analisar o impacto da heterogeneidade em modelos de duração com parâmetro de escala. Essas aproximações constituirão a base para a elaboração de um teste à detecção da heterogeneidade negligenciada. Será proposto um estimador GMM para a correcção do enviesamento induzido pela heterogeneidade na condição de momentos, sendo o seu comportamento testado com recurso a simulações de Monte Carlo.

O trabalho encontra-se organizado em seis capítulos: conceitos básicos, efeitos da heterogeneidade negligenciada em modelos de duração, modelos com uma duração, modelos com múltiplas durações, simulações e conclusões.

No capítulo 2 serão definidos alguns dos conceitos básicos dos modelos de duração, tais como a definição das funções utilizadas e a apresentação do fenómeno da censura.

No capítulo 3, vamos analisar o efeito da heterogeneidade negligenciada nos modelos de duração. A análise será efectuada inicialmente para o modelo mistura, sendo de seguida alargada aos modelos de duração com parâmetro de escala.

No capítulo 4 vamos utilizar as aproximações assintóticas para construir um teste score para detectar a heterogeneidade negligenciada. Serão determinadas aproximações para o enviesamento dos estimadores quando é ignorada a heterogeneidade com aplicação aos modelos Weibull e Log-logistic. De seguida, vamos apresentar um estimador GMM para corrigir o enviesamento causado pela heterogeneidade na

condição de momentos. Este estudo irá ser alargado a modelos com múltiplas durações no capítulo 5.

No capítulo 6 serão efectuadas simulações de Monte Carlo para analisar o enviesamento dos estimadores causado pela heterogeneidade negligenciada, assim como para analisar o comportamento do estimador GMM proposto. Para tal, serão consideradas duas distribuições paramétricas, Weibull e Log-logistic.

## 2. Conceitos básicos

Neste capítulo vamos introduzir os principais conceitos referentes aos modelos de duração e derivação das respectivas funções. O capítulo está estruturado em duas secções:

A secção 2.1 vai dar a conhecer algumas das funções mais utilizadas nos modelos de duração, nomeadamente a função sobrevivência, a função *hazard* e a função *hazard* integrada, mostrando as relações existentes entre elas.

Na secção 2.2 introduz-se a censura, fenómeno bastante comum nos modelos de duração, com especial atenção ao caso da censura à direita do tipo I.

### 2.1 Modelos de duração

Seja  $T$  uma variável aleatória contínua não negativa que representa a duração de um determinado acontecimento ou o tempo de permanência de um indivíduo num determinado estado. Para cada indivíduo, a data de início ou de entrada num estado é sempre dada por  $T = 0$ , de forma a que a variável  $T$  seja independente da data de início do processo, apenas dependendo da sua duração.

A sua função distribuição,  $F(t) = P(T \leq t)$ , representa a probabilidade da duração ser no máximo igual a  $t$ . A função sobrevivência,  $S(t) = P(T > t)$ , é a probabilidade da duração ser superior a  $t$ , sendo dada por

$$S(t) = 1 - F(t) \tag{2.1}$$

A função densidade

$$f(t) = \lim_{h \rightarrow 0} \frac{P(t < T \leq t + h)}{h}$$

pode ser interpretada como a probabilidade instantânea do indivíduo sair do estado em que se encontra no momento  $t$ , sendo definida como a derivada da função distribuição ou o simétrico da derivada da função densidade:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \quad (2.2)$$

A função *hazard*,  $\lambda(t)$ , representa a taxa de saída de um indivíduo de um determinado estado no momento  $t$ , sabendo que se manteve nesse estado até esse momento, sendo dada por

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T < t + h | T \geq t)}{h}$$

Em termos das funções densidade e sobrevivência, a função *hazard* pode ser escrita como

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt} \quad (2.3)$$

É também possível definir a função sobrevivência em termos da função *hazard*. Assim,

$$S(t) = \exp[-Z(t)] \quad (2.4)$$

onde  $Z(t)$  é a função *hazard* integrada que é dada por

$$Z(t) = \int_0^t \lambda(u) du \quad (2.5)$$

A função densidade de probabilidade, em termos das funções *hazard* e sobre-

vivência, é dada pelo produto entre essas duas funções

$$f(t) = \lambda(t) \exp \left[ - \int_0^t \lambda(u) du \right] \quad (2.6)$$

de forma que  $\int_0^\infty f(t) = 1$ .

## 2.2 Censura

Um problema que surge nos modelos de duração prende-se com o facto da duração de determinado acontecimento não poder ser exactamente observada para todos os indivíduos em estudo. Nestas situações, em que não é possível observar a duração completa, diz-se que as observações são censuradas.

Estas situações de censura podem ocorrer por duas razões distintas:

- O momento em que o indivíduo entrou no estado não é conhecido, sabendo-se apenas que, no início da observação, o indivíduo já se encontrava nesse estado - censura à esquerda;
- O momento em que o indivíduo sai do estado não é conhecido devido ao final da observação se ter verificado antes desse momento - censura à direita. Neste caso, o final do período em análise pode ter acontecido porque se verificou um determinado número prédefinido de saídas do estado, sendo todas as outras observações censuradas - censura do tipo II - ou porque se atingiu o período de tempo definido para a análise do fenómeno, pelo que todas as observações que não terminaram até essa data serão censuradas - censura do tipo I.

Em econometria, a censura mais comum é a censura à direita do tipo I, pelo que apenas esta será considerada no nosso estudo.

Vamos agora derivar a função densidade em presença de observações censuradas.



Para tal, suponhamos que estão  $n$  indivíduos em análise, estando associado ao  $i$ -ésimo indivíduo um tempo de permanência no estado em causa,  $T_i$ , e um tempo de censura,  $C_i$ . Como o tempo de censura  $C_i$  é dado pela diferença entre o momento correspondente ao final das observações e o momento de entrada do indivíduo no estado, será independente do tempo de permanência mas dependente da entrada do indivíduo no estado. O tempo exacto da permanência do indivíduo no estado apenas é observado se  $T_i \leq C_i$ , pelo que

$$t_i = \min(T_i, C_i) \quad (2.7)$$

Sendo  $d_i$  uma variável *dummy* que toma o valor 1 quando uma observação não é censurada e 0 quando essa observação é censurada,

$$d_i = \begin{cases} 1 & \text{se } T_i \leq C_i \\ 0 & \text{se } T_i > C_i \end{cases} \quad (2.8)$$

a função densidade conjunta de  $t_i$  e  $d_i$  é dada por

$$f(t_i)^{d_i} S(t_i)^{1-d_i} \quad (2.9)$$

### 3. Heterogeneidade negligenciada em modelos de duração

Se efectuarmos a estimação de modelos de duração negligenciando a heterogeneidade, vamos obter, geralmente, estimadores enviesados, pelo que se torna necessário incluir o efeito da heterogeneidade no modelo a estimar. Para incluir esse efeito, considera-se que a heterogeneidade é a realização de uma variável aleatória que será incorporada no modelo de duração.

Geralmente, a variável aleatória que representa a heterogeneidade é incluída nos modelos de duração através da sua incorporação no parâmetro de escala. Como é nesse parâmetro que se encontram as variáveis explicativas, a heterogeneidade pode ser vista como um problema de omissão de regressores.

Em distribuições pertencentes à família das funções *hazard* proporcionais (Cox, 1972), como a Weibull, incluir a variável da heterogeneidade no parâmetro de escala equivale a incluir essa variável de forma multiplicativa na função *hazard*, dando origem ao modelo mistura (Lancaster, 1979 e Vaupel, Mantom and Stallard, 1979).

No modelo mistura, o efeito da heterogeneidade é conhecido, originando uma dependência negativa da função *hazard* (Lancaster, 1979). No entanto, em distribuições de duração com parâmetros de escala não pertencentes à família das funções *hazard* proporcionais, como o caso da log-logistic, a inclusão da heterogeneidade no modelo como fazendo parte do parâmetro de escala pode originar efeitos diferentes no comportamento da função *hazard*.

Nas secções seguintes vamos analisar o efeito da heterogeneidade negligenciada nos modelos de duração:

Na secção 3.1 será analisado o efeito da heterogeneidade quando é incluída de

forma multiplicativa na função *hazard*, quer pela análise da função *hazard* marginal, quer pela utilização de aproximações assintóticas;

Na secção 3.2, utilizando aproximações assintóticas, vamos analisar o efeito da heterogeneidade quando esta é incluída no parâmetro de escala;

O problema da heterogeneidade negligenciada será comparado ao erro de medida na variável dependente na secção 3.3.

### 3.1 Modelo mistura *hazard* proporcional

Neste modelo, a variável de heterogeneidade,  $u$ , é incluída multiplicativamente na função *hazard* e, para garantir que essa variável seja sempre positiva, considera-se que  $u = e^v$ . Seja a função *hazard* condicional em  $x$  e em  $v$  dada por

$$\lambda(t|x, v) = \lambda(t|x)e^v$$

onde  $\lambda(t|x)$  é uma função *hazard* pertencente à família das funções *hazard* proporcionais.

Como  $v$  não é observável e varia ao longo da população, na prática, não podemos utilizar este modelo para construir uma verosimilhança com os dados observados. Para ultrapassar esta situação, vamos assumir  $v$  como sendo a realização de uma variável aleatória  $V$  com domínio  $\Re$  e função distribuição  $H(v)$ , com  $E(V) = 0$  e  $\text{Var}(V) = \sigma^2$ , e que  $v$  é independente de  $x$  e de  $t$ . Este modelo é o modelo mistura *hazard* proporcional (mixture proportional hazard, MPH), sendo fácil constatar que à medida que  $v \rightarrow \infty$  verifica-se um aumento do valor da função *hazard*, ou seja, uma aceleração na taxa de saída do estado.

### 3.1.1 Análise utilizando a função *hazard* marginal

Com vista a proceder à estimação do modelo mistura e, como não é possível efectuar a sua estimação condicional na heterogeneidade, necessitamos de marginalizar a variável da heterogeneidade do modelo de forma a manter o seu efeito na função distribuição.

Para tal, vamos integrar o produto da função distribuição de  $T$  condicional em  $V$  pela função densidade de  $V$ ,  $h(v)$ , de forma a obter a função distribuição de  $T$  marginal em  $V$

$$F_m(t) = \int F(t|v)h(v)dv \quad (3.1)$$

onde, para simplificação da notação, foi omitida a variável  $x$ . Essa omissão será efectuada ao longo deste trabalho, pelo que todas as funções de  $T$  são condicionais em  $x$ .

Como em Lancaster (1990), vamos analisar o efeito da heterogeneidade na função *hazard*. Para tal, vamos considerar a função sobrevivência do modelo mistura que é dada por

$$S_m(t) = \int S(t|v)h(v)dv = \int e^{-e^v Z(t)}h(v)dv \quad (3.2)$$

Utilizando (2.3), obtemos a função *hazard* mistura

$$\lambda_m(t) = \frac{\int \lambda(t|v)S(t|v)h(v)dv}{\int S(t|v)h(v)dv} = \frac{\lambda(t) \int e^v e^{-e^v Z(t)}h(v)dv}{\int e^{-e^v Z(t)}h(v)dv} \quad (3.3)$$

Consideremos a seguinte função densidade de  $V$  condicional em  $T \geq t$

$$p(e^v|T \geq t) = \frac{S(t|v)h(v)}{\int S(t|v)h(v)dv}$$

de forma a que

$$E(e^v|T \geq t) = \frac{\int e^v e^{-e^v Z(t)} h(v) dv}{\int e^{-e^v Z(t)} h(v) dv} \quad (3.4)$$

que representa a média da heterogeneidade nos indivíduos que ainda não saíram do estado.

Diferenciando (3.4) em ordem a  $t$  obtém-se

$$\begin{aligned} \frac{dE(e^v|T \geq t)}{dt} &= \lambda(t) \left[ \left( \frac{\int e^v e^{-e^v Z(t)} h(v) dv}{\int e^{-e^v Z(t)} h(v) dv} \right)^2 - \frac{\int e^{v^2} e^{-e^v Z(t)} h(v) dv}{\int e^{-e^v Z(t)} h(v) dv} \right] = \\ &= -\lambda(t) \left[ E(e^{v^2}|T \geq t) - E^2(e^v|T \geq t) \right] = \\ &= -\lambda(t) \text{Var}(e^v|T \geq t) \end{aligned} \quad (3.5)$$

Considerando (3.3) e (3.4), podemos escrever a função *hazard* mistura como

$$\lambda_m(t) = \lambda(t) E(e^v|T \geq t)$$

Derivando em ordem a  $t$  e considerando (3.5) obtemos

$$\frac{d\lambda_m(t)}{dt} = \frac{d\lambda(t)}{dt} E(e^v|T \geq t) - \lambda(t)^2 \text{Var}(e^v|T \geq t)$$

que após dividir por  $\lambda_m(t)$  vem

$$\frac{d \log \lambda_m(t)}{dt} = \frac{d \log \lambda(t)}{dt} - \lambda(t) \frac{\text{Var}(e^v|T \geq t)}{E(e^v|T \geq t)} \quad (3.6)$$

que representa a taxa de variação proporcional da função mistura *hazard* proporcional.

No segundo membro da equação (3.6), o primeiro termo representa a variação proporcional da função *hazard* quando é ignorada a heterogeneidade e, como o segundo termo é sempre positivo, verifica-se que  $\frac{d \log \lambda_m(t)}{dt} < \frac{d \log \lambda(t)}{dt}$ , ou seja, a função *hazard*

marginal em  $V$ , em percentagem, cai mais rapidamente ou cresce mais lentamente que a função *hazard* que ignora a heterogeneidade. Assim, a estimação do modelo negligenciando a heterogeneidade vai originar o enviesamento dos parâmetros funcionais da distribuição de  $T$ .

Por outro lado, apesar de  $E(V)$  não depender de  $x$ ,  $E(V, T \geq t)$  depende de  $x$  devido à presença de  $S(t|v)$  em  $p(e^v|T \geq t)$ . Como consequência, o efeito das variáveis explicativas na função *hazard* em que é negligenciada a heterogeneidade vai ser diferente do seu efeito na função *hazard* marginal em  $V$ , pelo que a estimação do modelo ignorando a heterogeneidade vai originar enviesamento nos regressores (Lancaster, 1990).

### 3.1.2 Análise utilizando aproximações assintóticas

Uma outra forma de analisar os efeitos da heterogeneidade negligenciada na função *hazard*, passa por considerar que a variância da heterogeneidade é pequena e utilizar aproximações assintóticas para calcular uma aproximação ao modelo com heterogeneidade (Lancaster, 1990 e Chesher, Dumangane e Smith, 2002).

Tal como anteriormente, consideremos que a heterogeneidade é dada por  $e^v$ , tendo  $V$  uma função distribuição  $H(v)$  tal que  $E(V) = 0$  e  $\text{Var}(V) = \sigma^2$ .

Seja a função distribuição dada por  $F(t, e^v)$ . Fazendo a aproximação em série de Taylor até à segunda ordem em torno de  $v = 0$  e ignorando os termos de ordem superior a dois, obtém-se

$$F(t, e^v) \simeq F(t, 1) + v F^v(t, 1) + \frac{v^2}{2}(F^{vv}(t, 1) + F^{vv}(t, 1))$$

onde  $F^v(t, 1)$  e  $F^{vv}(t, 1)$  correspondem à primeira e segunda derivada de  $F(t, 1)$  em

ordem a  $v$  e " $\simeq$ " representa uma aproximação de ordem  $O(\sigma^2)$  onde  $\lim_{\sigma \rightarrow 0} \frac{O(\sigma^2)}{\sigma^2} = 0^1$ .

Usando as hipóteses sobre os momentos de  $V$ , obtém-se a aproximação da função distribuição

$$F_a(t, e^v) = F(t, 1) + \frac{\sigma^2}{2}(F^v(t, 1) + F^{vv}(t, 1)) \quad (3.7)$$

onde, deste ponto em diante, a letra  $a$  em índice quer dizer aproximação.

Se a heterogeneidade entrar multiplicativamente na função *hazard*, a função distribuição com heterogeneidade é dada por

$$F(t, e^v) = 1 - \exp(-Z(t)e^v) \quad (3.8)$$

onde  $Z(t)$  corresponde à função *hazard* integrada.

Calculando em ambos os membros de (3.8) a primeira e segunda derivada em ordem a  $v$  obtemos respectivamente

$$e^v F^v(t, e^v) = Z(t)e^v \exp(-Z(t)e^v)$$

e

$$e^v F^v(t, e^v) + e^{2v} F^{vv}(t, e^v) = \exp(-Z(t)e^v)(Z(t)e^v - Z(t)^2 e^{2v})$$

que avaliando em  $v = 0$  vem

$$F^v(t, 1) = -Z(t) \exp(-Z(t))$$

e

$$F^v(t, 1) + F^{vv}(t, 1) = \exp(-Z(t))(Z(t) - Z(t)^2)$$

---

<sup>1</sup> Para que esta aproximação seja válida são condições suficientes (a) existir terceiro momento de  $V$  e (b) a função distribuição ter terceira derivada finita.

Sendo  $F(t, 1) = F(t)$ , podemos escrever a aproximação da função distribuição como

$$F_a(t, e^v) = F(t) + \frac{\sigma^2}{2} \exp(-Z(t))(Z(t) - Z(t)^2)$$

Considerando (2.1) e (2.4), a aproximação da função sobrevivência, tal como em Chesher, Dumangane e Smith (2002), é dada por

$$S_a(t) = S(t) \left( 1 + \frac{\sigma^2}{2} (Z(t)^2 - Z(t)) \right) \quad (3.9)$$

que tem as características dessa função, nomeadamente  $S_a(0) = 1$  e  $S_a(\infty) = 0$  e, quando  $\sigma^2 = 0$  verifica-se que  $S_a(t) = S(t)$ .

Calculando  $-\frac{d \log S_a(t)}{dt}$  obtém-se a aproximação da função *hazard*

$$\lambda_a(t) = \lambda(t) \left( 1 + \frac{\sigma^2}{2} (1 - 2Z(t)) \right) \quad (3.10)$$

Diferenciando o logaritmo da aproximação da função *hazard* em ordem a  $t$ , obtém-se a variação proporcional da aproximação da função *hazard*

$$\frac{d \log \lambda_a(t)}{dt} = \frac{d \log \lambda(t)}{dt} - \sigma^2 \lambda(t) \quad (3.11)$$

podendo-se tirar as mesmas conclusões do caso anterior, ou seja, como o segundo termo do segundo membro é sempre positivo e, sendo o primeiro termo do segundo membro a variação proporcional da função *hazard* onde é negligenciada a heterogeneidade, a função *hazard* com heterogeneidade, em percentagem, cai mais rapidamente ou cresce mais lentamente que a função *hazard* onde a heterogeneidade é ignorada.



## 3.2 Modelo com heterogeneidade no parâmetro de escala

Vamos agora analisar o impacto da heterogeneidade quando esta não é incluída de forma multiplicativa na função *hazard*, mas como fazendo parte de um parâmetro de escala. Como é no parâmetro de escala que se incluem as variáveis explicativas do modelo, a heterogeneidade negligenciada poderá, neste caso, ser interpretada como uma omissão de variáveis.

**Teorema 1** *A variável aleatória  $Y$  tem função distribuição com parâmetro de escala  $\theta$ , se, para todo e qualquer  $y$  e  $\theta$ ,  $F(y, \theta) = F(y \theta)$ .*

São exemplos de funções distribuição com parâmetro de escala as distribuições Weibull e Log-logistic.

Se o modelo paramétrico pertencer à família de distribuições com parâmetro de escala e a heterogeneidade for introduzida no modelo através desse parâmetro, então

$$F(t, \theta e^v) = F(t \theta e^v) \quad (3.12)$$

Tal como foi efectuado na secção anterior, vamos derivar ambos os membros de (3.12) em ordem a  $v$ . A primeira e segunda derivadas serão dadas respectivamente por<sup>2</sup>

$$\theta e^v F^v(t, \theta e^v) = \theta e^v t F'(t \theta e^v)$$

e por

$$\theta e^v F^v(t, \theta e^v) + \theta^2 e^{2v} F^{vv}(t, \theta e^v) = \theta e^v t F'(t \theta e^v) + \theta^2 e^{2v} t^2 F''(t \theta e^v)$$

---

<sup>2</sup> Deste ponto em diante, ' e '' representam a primeira e segunda derivada da função em ordem a  $t$ .

que após avaliar em  $v = 0$  se obtém

$$F^v(t, \theta) = t F'(t \theta)$$

e

$$F^v(t, \theta) + F^{vv}(t, \theta) = t F'(t \theta) + t^2 F''(t \theta)$$

Utilizando o resultado obtido em (3.7), a aproximação da função distribuição pode ser escrita como<sup>3</sup>

$$F_a(t, e^v) = F(t) + \frac{\sigma^2}{2}(t F'(t) + t^2 F''(t)) \quad (3.13)$$

Sendo  $F'(t) = -S'(t)$  e  $F''(t) = -S''(t)$  e considerando (2.1), a aproximação da função sobrevivência é dada por

$$\begin{aligned} S_a(t) &= S(t) + \frac{\sigma^2}{2}(t S'(t) + t^2 S''(t)) \\ &= S(t) \left( 1 + \frac{\sigma^2}{2} A(t) \right) \end{aligned} \quad (3.14)$$

onde

$$\begin{aligned} A(t) &= t \frac{S'(t)}{S(t)} + t^2 \frac{S''(t)}{S(t)} \\ &= -t Z'(t) + t^2 Z'(t)^2 - t^2 Z''(t) \\ &= -t \lambda(t) + t^2 \lambda(t) - t^2 \lambda'(t) \end{aligned}$$

Para obtermos a aproximação da função *hazard*, vamos calcular  $-\frac{d \log S_a(t)}{dt}$ :

$$\lambda_a(t) = \lambda(t) + \frac{\sigma^2}{2} B(t) \quad (3.15)$$

---

<sup>3</sup> Para simplificação da notação, na função  $F(t, \theta)$  foi omitido o parâmetro de escala  $\theta$ , sendo escrita como  $F(t)$

onde

$$B(t) = t^2 \lambda''(t) + 3t \lambda'(t) - 2t^2 \lambda'(t) \lambda(t) - 2t \lambda(t)^2 + \lambda(t)$$

sendo estes resultados equivalentes aos obtidos em Chesher, Dumangane e Smith (2002) para o problema de erro de medida multiplicativo na variável dependente.

Pode-se verificar que o impacto da heterogeneidade não é tão simples como no modelo mistura, podendo mesmo ter diferentes impactos na função *hazard*, dependendo da distribuição da duração e do valor da variável  $T$ . Apenas para alguns casos se verifica um efeito sempre constante e equivalente ao verificado para o modelo mistura. É o caso das distribuições em que  $\lambda(t, \theta) = \theta \lambda(t)$ , como é exemplo a distribuição Weibull. Nesta distribuição em particular, a derivada do logaritmo da aproximação da função *hazard* é dada por  $\frac{d \log(\lambda(t))}{dt} - \sigma^2 \lambda(t)$ , que é o mesmo valor que foi encontrado na secção anterior para o modelo mistura.

### 3.3 Heterogeneidade e erro de medida

A heterogeneidade negligenciada, quando é introduzida como fazendo parte do parâmetro de escala, pode ser comparada ao problema do erro de medida na variável dependente (Lancaster, 1990 e Chesher, Dumangane e Smith, 2002). No entanto, enquanto que a distribuição da heterogeneidade é independente dos parâmetros da distribuição de  $T$ , reescrever o problema da heterogeneidade como um problema de erro de medida da variável dependente origina que a distribuição desse erro de medida seja dependente de parâmetros da distribuição de  $T$ .

Consideremos que  $T$  tem uma distribuição com parâmetro de escala e com *shape parameter*  $\alpha$ , com  $\alpha > 0$ , que entra na função distribuição como expoente da variável

$t$ , como é o caso das distribuições Weibull e Log-logistic, onde temos  $F(t, \alpha) = F(t^\alpha)$ .

Neste caso, a função distribuição com heterogeneidade é dada por

$$F(t, \theta, e^v, \alpha) = F(t^\alpha e^v \theta) = F(t e^u, \theta, \alpha)$$

onde  $\theta$  é o parâmetro de escala e  $u = \frac{v}{\alpha}$ , sendo a distribuição de  $U$  tal que  $E(U) = 0$  e  $\text{Var}(U) = \frac{\sigma^2}{\alpha^2}$ , o que origina que a distribuição de  $U$  está dependente do valor do *shape parameter*  $\alpha$ .

Nestes modelos, a aproximação da função distribuição dada em (3.13) é reescrita como

$$F_a(t, e^v) = F(t) + \frac{\sigma^2}{2\alpha^2} (t F'(t) + t^2 F''(t)) \quad (3.16)$$

A aproximação da função densidade é obtida pela derivada da função (3.16), sendo dada por

$$f_a(t) = f(t) + \frac{\sigma^2}{2\alpha^2} (f(t) + 3t f'(t) + t^2 f''(t)) \quad (3.17)$$

Estas aproximações são semelhantes às aproximações obtidas em Chesher, Dummangane e Smith (2002), com  $\sigma^2$  substituído por  $\frac{\sigma^2}{\alpha^2}$ .

## 4. Modelos com uma duração

Como referido anteriormente, um dos métodos muito utilizado para a estimação de modelos com heterogeneidade negligenciada consiste em assumir uma distribuição paramétrica para a heterogeneidade. No entanto, se assumirmos uma forma incorrecta para a distribuição da heterogeneidade, mesmo que a forma funcional do modelo de duração esteja correcta, a estimação paramétrica poderá produzir estimadores enviesados (Heckman e Taber, 1984).

Na nossa análise, que será efectuada a modelos de duração com *shape parameter*, iremos considerar que a heterogeneidade é introduzida no modelo como parte integrante do parâmetro de escala.

Na secção 4.1, utilizando as aproximações assintóticas derivadas no capítulo anterior e, assumindo que a dispersão da heterogeneidade é pequena, vamos derivar uma estatística de teste baseada no teste score para detecção da heterogeneidade.

Na secção 4.2 serão apresentadas aproximações para a inconsistência do estimador GMM quando é ignorada a heterogeneidade na estimação do modelo.

Na secção 4.3, vamos propor um estimador GMM que corrige o enviesamento induzido pela heterogeneidade na condição de momentos.

### 4.1 Teste à heterogeneidade negligenciada

Tal como se verificou anteriormente, ignorar a heterogeneidade na análise de modelos de duração produz resultados inconsistentes em relação à forma como a função *hazard* depende da duração. O efeito das variáveis explicativas no modelo

com heterogeneidade é, geralmente, diferente do efeito dessas mesmas variáveis no modelo em que se ignora a sua presença. Assim, a estimação do modelo por Máxima Verosimilhança, mesmo que a forma assumida para a distribuição da duração seja a correcta, mas ignorando a presença da heterogeneidade, pode originar estimadores inconsistentes.

Como tal, é conveniente efectuar um teste à presença de heterogeneidade. O teste aqui proposto é um teste score à hipótese da variância da heterogeneidade ser nula,  $H_0 : \sigma^2 = 0$ . Este teste, que pertence à classe de testes da Matriz de Informação (ver Chesher, 1984), foi aplicado por Chesher, Dumangane e Smith (2002) ao contexto específico dos modelos de duração com erro de medida na variável dependente.

A aproximação à função densidade obtida em (3.17) é uma aproximação à função densidade do verdadeiro modelo para pequena dispersão da heterogeneidade. Como esta função é sensível à heterogeneidade, vai constituir a base para a construção da função verosimilhança aproximada<sup>4</sup> para os parâmetros de  $f(t)$  e  $\sigma^2$ . O logaritmo da sua função verosimilhança aproximada é dado por

$$L_a(\phi, \sigma^2) = \log \left[ f(t) + \frac{\sigma^2}{2\alpha^2} (f(t) + 3t f'(t) + t^2 f''(t)) \right] \quad (4.1)$$

onde  $\phi$  é o conjunto dos  $p$  parâmetros de  $f(t)$  que se pretendem estimar.

A contribuição para o score de  $\sigma^2$  no ponto  $\sigma^2 = 0$  é dada por  $D_{\sigma^2}(t)$ , derivada de (4.1) em ordem a  $\sigma^2$  avaliada em  $\sigma^2 = 0$

$$\begin{aligned} D_{\sigma^2}(t) &= (f(t) + 3t f'(t) + t^2 f''(t)) \frac{1}{\alpha^2 f(t)} \\ &= \frac{1}{\alpha^2} \left( 1 + 3t \tilde{f}'(t) + \left( t^2 \tilde{f}''(t) + \tilde{f}'(t)^2 \right) \right) \end{aligned} \quad (4.2)$$

<sup>4</sup> Não é uma função verosimilhança porque a aproximação da função densidade dada em (3.17) não tem as propriedades de uma função densidade.

onde  $\tilde{f}(t) = \log f(t)$ .

A variância assintótica de (4.2),  $\Omega$ , necessária para construir a estatística de teste, é dada por

$$\Omega = E[D_{\sigma^2}(t)^2] - E[D_{\sigma^2}(t)D_{\phi}(t)]' E[D_{\phi}(t)D_{\phi}(t)]^{-1} E[D_{\sigma^2}(t)D_{\phi}(t)] \quad (4.3)$$

onde  $E[.]$  representa o valor esperado em relação à distribuição sem heterogeneidade e  $D_{\phi}(t)$  representa a derivada de  $\log f(t)$  em ordem a  $\phi$ .

Seja

$$\hat{D}_{\sigma^2}^n = n^{-1/2} \sum_{i=1}^n \hat{D}_{\sigma^2}(t_i)$$

onde  $\hat{D}_{\sigma^2}(t_i)$  é igual a  $D_{\sigma^2}(t)$  com os parâmetros desconhecidos  $\phi$  substituídos pelos seus estimadores de Máxima Verosimilhança,  $\hat{\phi}$ , obtidos sob a hipótese nula.

A estatística do teste score para  $H_0 : \sigma^2 = 0$  é dada por

$$T = \frac{\hat{D}_{\sigma^2}^n{}^2}{\hat{\Omega}} \sim \chi_{(1)}^2 \quad (4.4)$$

onde  $\hat{\Omega}$  é o estimador de  $\Omega$ .

Dois estimadores de  $\Omega$  podem ser considerados: a forma eficiente, EFF, na qual os parâmetros desconhecidos,  $\phi$ , em (4.3) são substituídos pelos seus estimadores de Máxima Verosimilhança obtidos sob a hipótese nula e a forma *outer-product-of-gradients*, OPG, na qual os valores esperados em (4.3) são substituídos pelas médias amostrais dos valores obtidos pela estimação por Máxima Verosimilhança de  $\phi$  sob a hipótese nula.

Chesher, Dumangane e Smith (2002) mostraram que este teste pertence à classe dos testes da Matriz de Informação e, com simulações de Monte Carlo, usando aproximações assintóticas de segunda ordem, demonstraram que nos casos estudados,

dos estimadores de  $\Omega$ , a forma eficiente, EFF, produz melhores resultados ao nível da dimensão e potência do teste que a forma OPG. Apesar dos resultados obtidos serem para o teste ao erro de medida multiplicativo na variável dependente, também mostraram que esse teste é equivalente ao teste para modelos com heterogeneidade incluída no parâmetro de escala.

## 4.2 Enviesamento dos estimadores

Como constatamos, negligenciar a heterogeneidade na estimação de modelos de duração produz estimadores enviesados. Nesta secção, seguindo o método utilizado por Dumangane (2001) para o erro de medida na variável dependente, vamos analisar a inconsistência dos estimadores GMM quando se ignora a heterogeneidade e derivar uma expressão para a aproximação do limite em probabilidade do estimador de Máxima Verosimilhança.

### 4.2.1 Observações completas

Consideremos que a verdadeira distribuição da duração é gerada pela densidade  $f(t, \phi, \sigma^2)$ , que depende do  $(p \times 1)$  vector de parâmetros  $\phi$  e do escalar  $\sigma^2$ , a variância da heterogeneidade negligenciada. Desta forma, as durações observadas são uma realização da densidade para  $\theta_0 = \{\phi_0, \sigma_0^2\}$  que é o  $((p + 1) \times 1)$  vector dos verdadeiros parâmetros.

Se a distribuição da duração for conhecida e tivermos em conta a heterogeneidade, a estimação por Máxima Verosimilhança irá produzir estimadores consistentes do vector dos parâmetros de interesse. No entanto, como a heterogeneidade não



é observada, o analista não a vai considerar na estimação, pelo que vai estimar  $f(t, \phi) = f(t, \phi, 0)$ , que não depende do parâmetro da heterogeneidade, considerando que  $\sigma^2 = 0$ .

Como já foi referido, negligenciar a heterogeneidade vai alterar a forma funcional da distribuição da duração, pelo que a estimação de  $f(t, \phi)$  vai produzir estimadores enviesados dos parâmetros da distribuição.

Suponhamos que o modelo de duração sem heterogeneidade é caracterizado pela seguinte condição de momentos

$$E_{\phi}[g(T, \phi_0)] = 0 \quad (4.5)$$

onde  $g(t, \phi)$  é o vector  $(q \times 1)$  definido como função de  $\phi$ , com  $q \geq p$ , condicional em  $x$  e onde  $E_{\phi}[\cdot]$  representa o valor esperado em relação à distribuição sem heterogeneidade, caracterizada pelo vector de parâmetros  $\phi$ .

O estimador GMM é definido como o maximizante do simétrico do quadrado da distância Euclidiana da média amostral:

$$\hat{\phi}_n : \arg \max_{\phi} \hat{Q}_n(\phi) = -\hat{g}_n(\phi)' \hat{W} \hat{g}_n(\phi) \quad (4.6)$$

sendo  $\hat{g}_n(\phi) = n^{-1} \sum_{i=1}^n g(t_i, \phi)$  e  $\hat{W}$  uma matriz  $(q \times q)$  semi-definida positiva que depende das observações apenas através de  $\phi$ .

Como a presença da heterogeneidade altera a forma funcional da distribuição da duração, a condição de momentos não vai ser satisfeita, ou seja,  $E_{\theta_0}[g(t, \phi)] \neq 0$ , onde  $E_{\theta_0}[\cdot]$  representa o valor esperado em relação à distribuição com heterogeneidade, a verdadeira distribuição, originando o enviesamento dos estimadores dos parâmetros de interesse.

Seja  $g_0(\phi) = E_{\theta_0}[g(t, \phi)]$ ,  $G_0(\phi) = E_{\theta_0}[\nabla_{\phi} g(t, \phi)]$  e  $Q_0(\phi) = -g_0'(\phi) W g_0(\phi)$ .

O limite em probabilidade do estimador,  $\tilde{\phi}(\theta_0)$ , é a solução implícita do sistema de  $(q \times 1)$  equações,

$$\tilde{\phi}(\theta_0) : \quad G_0(\tilde{\phi}(\theta_0))' W g_0(\tilde{\phi}(\theta_0)) = 0 \quad (4.7)$$

que escrito na forma de integrais vem

$$E_X \left[ \left[ \int_0^\infty \nabla_\phi g(t, \tilde{\phi}(\theta_0))' f(t, \theta_0) dt \right] W \left[ \int_0^\infty g(t, \tilde{\phi}(\theta_0)) f(t, \theta_0) dt \right] \right] = 0 \quad (4.8)$$

onde  $E_X[\cdot]$  representa o valor esperado condicional em  $x$ .

Tal como efectuado por Kiefer e Skoog (1984), Chesher, Lancaster e Irish (1985) e Dumangane (2001), podemos obter uma aproximação do limite em probabilidade do estimador se expandirmos  $\tilde{\phi}(\theta_0)$  em série de Taylor em torno de  $\sigma_0^2 = 0$ . Considerando que  $\tilde{\phi}(\phi_0, 0) = \phi_0$ , a aproximação do limite em probabilidade do estimador que ignora a heterogeneidade é dada por

$$\tilde{\phi}(\theta_0) \simeq \phi_0 + \left. \frac{d\tilde{\phi}(\theta_0)}{d\sigma_0^2} \right|_{\sigma_0^2=0} \sigma_0^2 \quad (4.9)$$

No entanto, é necessário calcular  $\frac{d\tilde{\phi}(\theta_0)}{d\sigma_0^2}$ . Para tal, vamos diferenciar a equação (4.8) em ordem a  $\tilde{\phi}(\theta_0)$  e a  $\sigma_0^2$ , pelo que necessitamos de calcular as suas derivadas parciais. Seja  $P(t, \tilde{\phi}, \theta_0)$  igual ao primeiro membro de (4.8), onde  $f(t, \theta_0)$  foi substituída pela sua aproximação

$$f_a(t, \theta_0) = f(t, \phi_0) + \frac{\sigma^2}{2\alpha^2} h(t, \phi_0)$$

onde

$$h(t, \phi_0) = f(t, \phi_0) + 3t f'(t, \phi_0) + t^2 f''(t, \phi_0)$$

As derivadas parciais em ordem a  $\tilde{\phi}$  e a  $\sigma_0^2$  são dadas respectivamente por

$P_\phi(t, \tilde{\phi}, \theta_0)$  e por  $P_{\sigma^2}(t, \tilde{\phi}, \theta_0)$ :

$$P_\phi(t, \tilde{\phi}, \theta_0) = \left[ \int_0^\infty \nabla_\phi g(t, \tilde{\phi}(\theta_0))' f(t, \theta_0) dt \right] W \left[ \int_0^\infty g(t, \tilde{\phi}(\theta_0)) f(t, \theta_0) dt \right] + \\ + \left[ \int_0^\infty \nabla_\phi g(t, \tilde{\phi}(\theta_0))' f(t, \theta_0) dt \right] W \left[ \int_0^\infty \nabla_\phi g(t, \tilde{\phi}(\theta_0)) f(t, \theta_0) dt \right]$$

$$P_{\sigma^2}(t, \tilde{\phi}, \theta_0) = \left[ \int_0^\infty \nabla_\phi g(t, \tilde{\phi}(\theta_0))' \frac{h(t, \phi_0)}{2\alpha^2} dt \right] W \left[ \int_0^\infty g(t, \tilde{\phi}(\theta_0)) f(t, \theta_0) dt \right] + \\ + \left[ \int_0^\infty \nabla_\phi g(t, \tilde{\phi}(\theta_0))' f(t, \theta_0) dt \right] W \left[ \int_0^\infty g(t, \tilde{\phi}(\theta_0)) \frac{h(t, \phi_0)}{2\alpha^2} dt \right]$$

tendo sido omitidos os valores esperados de  $x$  para simplificação das notações.

Avaliando estas funções em  $\sigma^2 = 0$ , tendo em conta que  $\tilde{\phi}(\phi_0, 0) = \phi_0$  e que  $\int_0^\infty g(t, \tilde{\phi}(\theta_0)) f(t, \theta_0) dt = 0$  e aplicando o teorema da função implícita, obtemos a equação

$$\left( \left[ \int_0^\infty \nabla_\phi g(t, \tilde{\phi}(\theta_0))' f(t, \theta_0) dt \right] W \left[ \int_0^\infty \nabla_\phi g(t, \tilde{\phi}(\theta_0)) f(t, \theta_0) dt \right] \right) d\tilde{\phi}(\theta_0) + \\ + \left( \left[ \int_0^\infty \nabla_\phi g(t, \tilde{\phi}(\theta_0))' f(t, \theta_0) dt \right] W \left[ \int_0^\infty g(t, \tilde{\phi}(\theta_0)) \frac{h(t, \phi_0)}{2\alpha^2} dt \right] \right) d\sigma^2 = 0 \quad (4.10)$$

Sendo a matriz  $G(\phi)$  de dimensão  $(q \times p)$  dada por

$$G(\phi) = E_X[E_\phi[\nabla_\phi g(t, \phi)]] \quad (4.11)$$

e o vector  $(q \times 1)$  da aproximação da função enviesamento por  $\sigma^2 b(\phi) = E_X[E_\phi[g(t, \phi)]]$ , a aproximação da função enviesamento condicional, que representa o enviesamento na condição de momentos originado por ignorar a heterogeneidade, pode ser dada

por

$$\sigma^2 b(\phi) = \frac{\sigma^2}{2\alpha^2} \int_0^\infty g(t, \phi) h(t, \phi) dt \quad (4.12)$$

Seja  $G(\phi_0) = G_0$ . Usando a condição de momentos definida em (4.5) e re-escrevendo a equação (4.10) obtemos

$$\frac{d\tilde{\phi}(\theta_0)}{d\sigma_0^2} = -(G_0' W G_0)^{-1} G_0' W b(\phi_0)$$

pelo que a aproximação do limite em probabilidade do estimador GMM que ignora a heterogeneidade pode ser dada por

$$\tilde{\phi}(\theta_0) \simeq \phi_0 - \sigma_0^2 (G_0' W G_0)^{-1} G_0' W b(\phi_0) \quad (4.13)$$

Quando  $q = p$ , a matriz  $G_0$  é triangular e a expressão (4.13) pode ser simplificada para

$$\tilde{\phi}(\theta_0) \simeq \phi_0 - \sigma_0^2 G_0^{-1} b(\phi_0) \quad (4.14)$$

Se o modelo for definido parametricamente,  $\tilde{\phi}(\theta_0)$  é a aproximação do limite em probabilidade do estimador de Máxima Verosimilhança. Neste caso, a matriz dada por  $G(\phi_0) = -E_\phi[\nabla_{\phi\phi} \log f(t, \phi_0)]$  é a Matriz de Informação,  $I(\phi_0)$ , e  $\sigma_0^2 b(\phi_0)$  é a aproximação do enviesamento na equação score  $\nabla_\phi \log f(t, \phi_0)$  induzida pela heterogeneidade.

#### 4.2.2 Observações censuradas

Consideremos agora um modelo em que existem observações censuradas à direita. Seja  $d_i$  a variável indicadora da existência ou não de censura na  $i$ -ésima observação

de forma que  $d_i = 1$  se a  $i$ -ésima observação for completa e  $d_i = 0$  caso contrário, tal como definido em (2.8). A condição de momentos é dada por

$$E_\phi[d g_1(T, \phi) + (1 - d)g_0(T, \phi)] = 0 \quad (4.15)$$

onde  $g_1(T, \phi) = g(T, \phi|d = 1)$  e  $g_0(T, \phi) = g(T, \phi|d = 0)$ . Caso o modelo seja definido parametricamente, estas funções serão dadas, respectivamente, por  $\nabla_\phi \log f(t, \phi)$  e por  $\nabla_\phi \log S(t, \phi)$ .

Desta forma, o estimador GMM maximiza a função objectivo (4.6), onde  $\hat{g}_n(\phi)$  é dado por  $n^{-1} \sum_{i=1}^n d_i g_1(t_i, \phi) + n^{-1} \sum_{i=1}^n (1 - d_i)g_0(t_i, \phi)$ , sendo o sistema de equações que define o limite em probabilidade do estimador, equivalente a (4.8), dado por

$$\tilde{\phi}(\theta_0) : \quad E_{\theta_0}[d g_1(T, \tilde{\phi}(\theta_0)) + (1 - d)g_0(T, \tilde{\phi}(\theta_0))] \quad (4.16)$$

Para determinar o limite em probabilidade vamos proceder como anteriormente. No entanto, é necessário especificar qual o mecanismo de censura. Como a entrada de cada indivíduo no estado é aleatória, também o tempo de censura é aleatório, estando apenas dependente do momento da entrada do indivíduo no estado em causa. Desta forma, o tempo de censura,  $C_i$ , é uma variável aleatória, independente dos indivíduos e do tempo de permanência de cada indivíduo no estado, com uma função densidade de probabilidade e uma função sobrevivência dadas por  $r(c)$  e  $R(c)$  respectivamente.

Tendo em conta a condição (2.8) estabelecida para a variável indicadora da censura,  $d_i$ , e que em caso de observações não censuradas  $C_i = T_i$ , podemos escrever a expressão (4.16) na forma de integral

$$E_{\theta_0}[g(T, \tilde{\phi}(\theta_0))] = \int_0^c g_1(t, \tilde{\phi}(\theta_0))f(t, \tilde{\phi}(\theta_0))dt + \int_c^\infty g_0(c, \tilde{\phi}(\theta_0))f(t, \tilde{\phi}(\theta_0))dt$$

onde o primeiro e segundo integral correspondem, respectivamente, às observações

não censuradas e às observações censuradas. Como em presença de observações censuradas  $\int_c^\infty f(t, \tilde{\phi}(\theta_0))dt = S(c, \tilde{\phi}(\theta_0))$ , a expressão anterior pode ser escrita como

$$E_{\theta_0}[g(T, \tilde{\phi}(\theta_0))] = \int_0^c g_1(t, \tilde{\phi}(\theta_0))f(t, \tilde{\phi}(\theta_0))dt + g_0(c, \tilde{\phi}(\theta_0))S(c, \tilde{\phi}(\theta_0)) \quad (4.17)$$

Utilizando as aproximações da função densidade, (3.17), e da função sobrevivência, (3.9), e aplicando, tal como anteriormente, o teorema da função implícita, obtemos a aproximação do limite em probabilidade equivalente a (4.13) com

$$\begin{aligned} G(\phi_0) &= \int_0^c \nabla_\phi g_1(t, \phi_0)f(t, \phi_0)dt + \nabla_\phi g_0(c, \phi_0)S(c, \phi_0) \\ b(\phi_0) &= \frac{1}{2\alpha^2} \left[ \int_0^c g_1(t, \phi_0)h(t, \phi_0)dt + g_0(c, \phi_0)H(c, \phi_0) \right] \end{aligned} \quad (4.18)$$

onde

$$\begin{aligned} h(t, \phi_0) &= f(t, \phi_0) + 3t f'(t, \phi_0) + t^2 f''(t, \phi_0) \\ H(c, \phi_0) &= c S'(c, \phi_0) + c^2 S''(c, \phi_0) \end{aligned}$$

### 4.2.3 Exemplos

Nesta secção, vamos apresentar alguns exemplos para mostrar em que medida é que as aproximações do limite em probabilidade podem ser importantes na análise do efeito da heterogeneidade negligenciada em distribuições paramétricas. A análise será efectuada considerando duas distribuições paramétricas, Weibull e Log-logistic.

#### Exemplo 1: modelo Weibull

Suponhamos que a verdadeira distribuição da duração é gerada por uma dis-

tribuição Weibull. Para esta distribuição, a função *hazard* condicional em  $x$  é dada por

$$\lambda(t|x, \alpha, \beta) = \alpha t^{\alpha-1} \exp(x'\beta), \quad \alpha > 0$$

onde o vector associado aos coeficientes está dividido em  $\beta = \{\beta_0, \beta_1\}$  e sendo  $\beta_0$  definido de forma que  $x$  tenha valor esperado nulo e matriz de covariância  $\Sigma$ .

A matriz informação tem inversa e, com os parâmetros ordenados por  $\alpha, \beta_0$  e  $\beta_1$ , é dada por

$$i^{-1} = G(\phi)^{-1} = \begin{bmatrix} \alpha^2 p & -\alpha m p & \alpha \beta_1 p \\ \dots & 1 + m^2 p & -\beta_1 m p \\ \dots & \dots & \Sigma^{-1} + \beta_1' \beta_1 p \end{bmatrix} \quad (4.19)$$

onde  $m = 1 + \psi(1) - \beta_0$  e  $p = 1/\psi'(1)$ .

Sendo

$$b(\phi) = \begin{bmatrix} -\frac{2+m}{2\alpha} \\ -\frac{1}{2} \\ 0 \end{bmatrix}$$

e aplicando (4.14), o enviesamento dos estimadores é dado por

$$\begin{bmatrix} \tilde{\alpha}(\theta_0) \\ \tilde{\beta}_0(\theta_0) \\ \tilde{\beta}_1(\theta_0) \end{bmatrix} \simeq \begin{bmatrix} \frac{\psi'(1) - \sigma^2}{\psi'(1)} \alpha \\ \beta_0 + \sigma^2 \frac{2m - \psi'(1)}{2\psi'(1)} \\ \frac{\psi'(1) - \sigma^2}{\psi'(1)} \beta_1 \end{bmatrix} \quad (4.20)$$

Através da análise de (4.20), facilmente se verifica que a heterogeneidade negligenciada origina o enviesamento dos coeficientes dos regressores e do *shape parameter*, sendo esse enviesamento sempre em direcção a zero.

Se medirmos o enviesamento proporcional no *shape parameter*, verificamos que este é dado por

$$\frac{\tilde{\alpha}(\theta_0)}{\alpha} \simeq \frac{\psi'(1)}{\psi'(1) + \sigma^2} \quad (4.21)$$

pelo que o enviesamento apenas depende de  $\sigma^2$ , sendo invariável em relação ao seu verdadeiro valor (Lancaster, 1990). Este resultado é equivalente ao rácio entre as variâncias das distribuições sem heterogeneidade e com heterogeneidade, medidas em escala logaritmica,  $\frac{Var(\log T)}{Var(\log TH)}$ <sup>5</sup>.

## Exemplo 2: modelo Log-logistic

Consideremos agora que a verdadeira distribuição da duração é gerada por uma distribuição Log-logistic. A função *hazard* condicional em  $x$  é dada por

$$\lambda(t|x, \alpha, \beta) = \frac{\alpha t^{\alpha-1} \exp(x'\beta)}{1 + t^\alpha \exp(x'\beta)}, \quad \alpha > 0$$

onde o vector associado aos coeficientes é definido da mesma forma que no exemplo anterior.

A matriz informação tem inversa e, com os parâmetros ordenados por  $\alpha$ ,  $\beta_0$  e  $\beta_1$ , é dada por

$$i^{-1} = G(\phi)^{-1} = \begin{bmatrix} \alpha^2 m & \alpha \beta_0 m & \alpha \beta_1 m \\ \dots & (\beta_0^2 + p)m & \beta_0 \beta_1 m \\ \dots & \dots & (\beta_1^2 \Sigma + p)m \Sigma^{-1} \end{bmatrix} \quad (4.22)$$

onde  $m = 3/(1 + 2\psi'(1))$  e  $p = 1 + 2\psi'(1)$ .

<sup>5</sup>  $Var(\log TH)$  corresponde à variância do logaritmo da duração com heterogeneidade e  $Var(\log T)$  à variância do logaritmo da duração sem heterogeneidade.



Sendo

$$b(\phi) = \begin{bmatrix} -\frac{1}{4\alpha} \\ 0 \\ 0 \end{bmatrix}$$

e aplicando (4.14), o enviesamento dos estimadores é dado por

$$\begin{bmatrix} \tilde{\alpha}(\theta_0) \\ \tilde{\beta}_0(\theta_0) \\ \tilde{\beta}_1(\theta_0) \end{bmatrix} \simeq \begin{bmatrix} \frac{1 + 2\psi'(1) - 3\sigma^2/4}{1 + 2\psi'(1)} \alpha \\ \frac{1 + 2\psi'(1) - 3\sigma^2/4}{1 + 2\psi'(1)} \beta_0 \\ \frac{1 + 2\psi'(1) - 3\sigma^2/4}{1 + 2\psi'(1)} \beta_1 \end{bmatrix} \quad (4.23)$$

Analisando (4.23), verifica-se que, tal como no exemplo anterior, a heterogeneidade negligenciada origina que o enviesamento dos coeficientes dos regressores e do *shape parameter* seja sempre em direcção a zero.

Se medirmos o enviesamento proporcional no *shape parameter*, este é dado por

$$\frac{\tilde{\alpha}(\theta_0)}{\alpha} \simeq \frac{1 + 2\psi'(1)}{1 + 2\psi'(1) + 3\sigma^2/4} \quad (4.24)$$

pelo que o enviesamento, tal como no modelo anterior, apenas depende de  $\sigma^2$ . No entanto, se compararmos este resultado com o resultado obtido para o modelo Weibull, (4.21), verificamos que o modelo Log-logistic é mais robusto à heterogeneidade negligenciada, pois a proporção de enviesamento é sempre inferior para qualquer valor de  $\sigma^2$ .

### 4.3 Estimador GMM corrigido

Na secção anterior foi demonstrado que os estimadores GMM são inconsistentes

na estimação de modelos onde é negligenciada a heterogeneidade. De seguida, vamos derivar um estimador GMM que corrige o enviesamento induzido pela heterogeneidade na condição de momentos.

Este estimador foi proposto por Chesher (2000) para o problema do erro de medida nos regressores e aplicado por Dumangane (2001) ao problema do erro de medida na variável dependente. A ideia base para a sua construção é que, se a condição de momentos não é satisfeita devido a um problema de erro nos regressores, essa condição de momentos pode ser corrigida aproximadamente pela introdução de uma função que caracterize a proporção do enviesamento causado pelo erro de medida nos regressores.

Consideremos a distribuição caracterizada pela condição de momentos definida em (4.5). Sob o efeito de erro de medida nos regressores e, sendo  $\sigma^2 b(\phi)$  a aproximação da função enviesamento condicional, a condição de momentos corrigida do enviesamento induzido pelo erro de medida é dada por  $E_\theta[g(T, \phi) - \sigma^2 b(\phi)] \simeq 0$ .

Como ignorar a heterogeneidade nos modelos de duração vai alterar a forma como a função *hazard* depende de  $T$ , temos de introduzir uma função de  $x$  e de  $T$  com vista a corrigir de forma "estrutural" as condições de momentos, sendo essa correcção dada por  $E_\theta[g(T, \phi) - \sigma^2 b(t, \phi)] \simeq 0$ .

No apêndice A é demonstrado que a expressão (4.12) é equivalente a

$$\sigma^2 b(\phi) = \frac{\sigma^2}{2\alpha^2} \int_0^\infty (t g'(t, \phi) + t^2 g''(t, \phi)) f(t, \phi) dt \quad (4.25)$$

Sendo  $\sigma^2 b(t, \phi)$  a aproximação da função enviesamento estrutural

$$\sigma^2 b(t, \phi) = \frac{\sigma^2}{2\alpha^2} (t g'(t, \phi) + t^2 g''(t, \phi)) \quad (4.26)$$

a condição de momentos corrigida é dada por

$$E_{\theta}[g^c(T, \theta)] = E_{\theta}[g(T, \phi) - \sigma^2 b(T, \phi)]$$

que irá satisfazer a condição  $E_{\theta}[g^c(T, \theta)] \simeq 0$ .

Em caso de existência de censura através do mecanismo definido na secção anterior, a aproximação da função enviesamento condicional é dada por

$$\sigma^2 b(c, \phi) = \frac{\sigma^2}{2\alpha^2} \left[ \int_0^c g_1(t, \phi_0) h(t, \phi_0) dt + g_0(c, \phi_0) H(c, \phi_0) \right] \quad (4.27)$$

No apêndice A é demonstrado que a aproximação da função enviesamento estrutural, na presença de censura, é dada por

$$b(t, c, \phi) = \frac{1}{2\alpha^2} d(t, g'_1(t, \phi) + t^2 g''_1(t, \phi)) + q(c, \phi) \quad (4.28)$$

onde

$$q(c, \phi) = (c g_1(c, \phi) - c^2 g'_1(c, \phi) + c g_0(c, \phi)) f(c, \phi) + (c^2 g_1(c) - c^2 g_0(c, \phi)) f'(c, \phi)$$

sendo a condição de momentos corrigida obtida da mesma forma que no modelo em que não existe censura.

Nas figuras 4.1 e 4.2 podemos ver o enviesamento nas condições de momentos antes e após ter sido efectuada a correcção proposta. Para a elaboração dos gráficos foram utilizadas duas distribuições paramétricas para a duração, Weibull e Log-logistic, e duas distribuições para a variável da heterogeneidade: Log-normal (painéis da esquerda) e Gamma (painéis da direita) . No eixo horizontal encontra-se a proporção da variância causada pela variância da heterogeneidade, ou seja, o rácio

$$1 - \frac{Var(\log T)}{Var(\log TH)}.$$

Figura 4.1: Enviesamento das condições de momentos sem correcção (pontos) e com correcção (tracejado) para a distribuição Weibull com distribuição da heterogeneidade Lognormal (painéis da esquerda) e Gamma (painéis da direita)

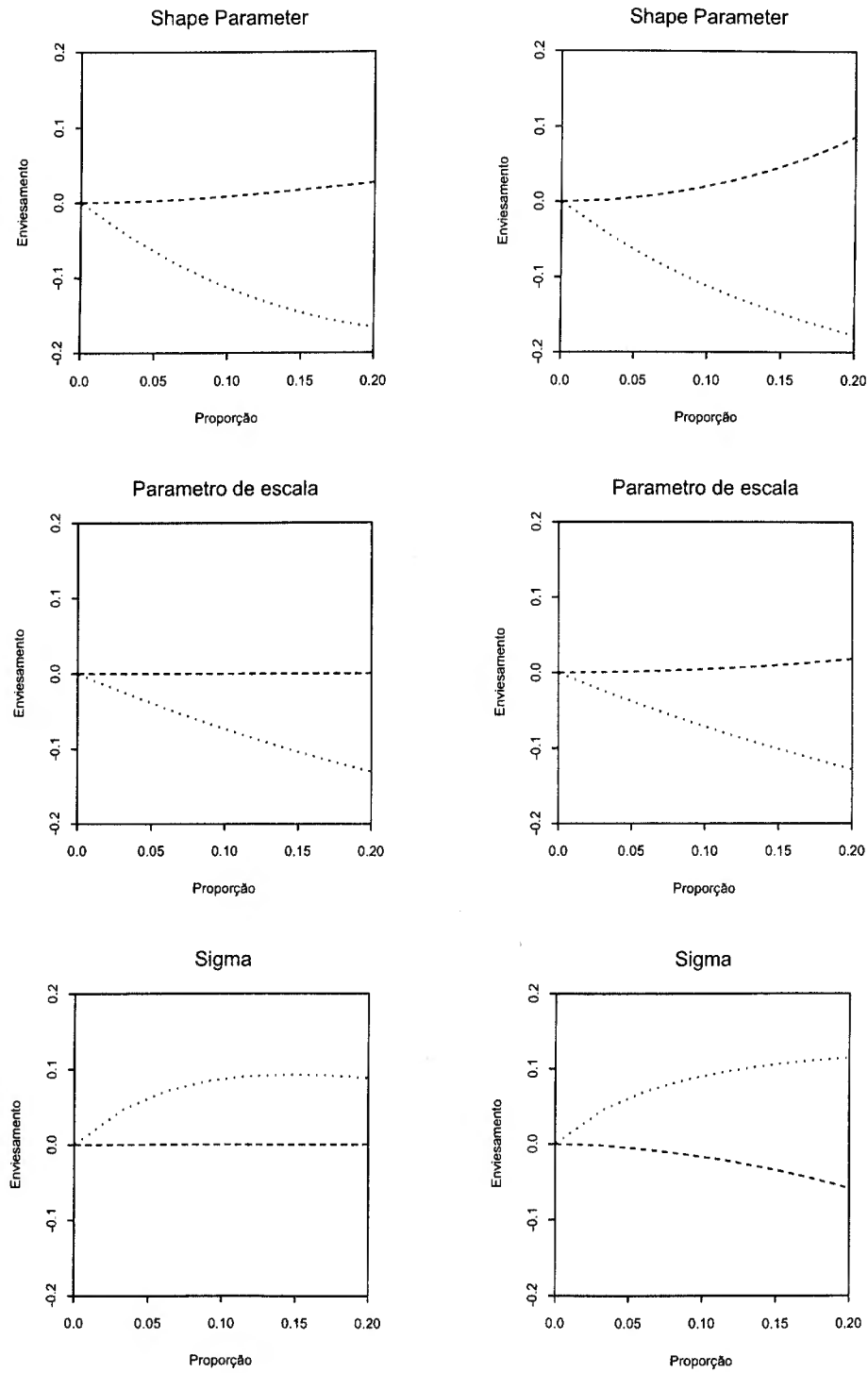
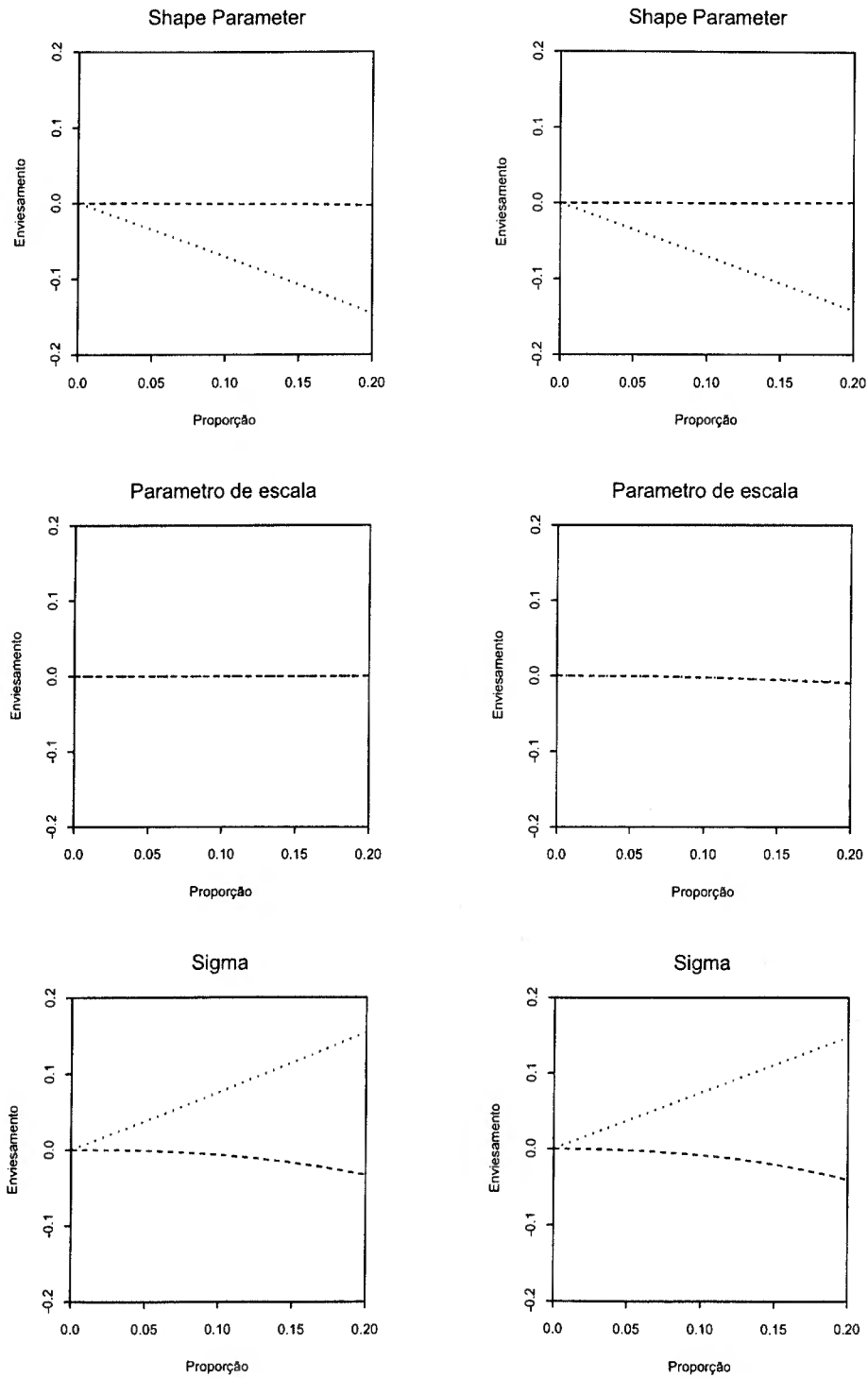


Figura 4.2: Enviesamento das condições de momentos sem correcção (pontos) e com correcção (tracejado) para a distribuição Log-logistic com distribuição da heterogeneidade Lognormal (painéis da esquerda) e Gamma (painéis da direita)



Através da figura 4.1, pode-se verificar que no modelo Weibull a condição de momentos sem correcção é muito sensível à heterogeneidade, quer para o *shape parameter* quer para o parâmetro de escala, apesar do enviesamento ser superior no *shape parameter*. Por outro lado, apesar de terem sido utilizadas duas distribuições para a heterogeneidade, as diferenças no enviesamento das condições de momentos não são significativas, sendo o enviesamento um pouco superior quando a heterogeneidade segue uma distribuição Gamma.

Quando utilizamos as condições de momentos corrigidas, tal como foi proposto, verifica-se que o enviesamento no parâmetro de escala é quase anulado na totalidade, principalmente quando a heterogeneidade segue uma distribuição Lognormal, enquanto que no *shape parameter* essa correcção é menor.

Na figura 4.2, utilizando o modelo Log-logistic, é fácil verificar que o enviesamento nas condições de momentos deve-se quase exclusivamente ao *shape parameter*. No entanto, ao utilizarmos a correcção da condição de momentos, o enviesamento é quase corrigido na sua totalidade até à proporção considerada.

No último gráfico de cada distribuição, encontra-se a condição de momentos para a variância da heterogeneidade,  $\sigma^2$ , sendo fácil constatar que, devido ao seu rápido enviesamento em ambas as distribuições, este parâmetro é facilmente detectado por um teste à heterogeneidade negligenciada baseado nesta condição de momentos<sup>6</sup>.

---

<sup>6</sup> Na elaboração do teste podemos nos deparar com problemas de identificação se a distribuição da heterogeneidade produzir um modelo mistura que não se distinga do modelo sem heterogeneidade.

## 5. Modelos com múltiplas durações

Neste capítulo vamos efectuar a análise a modelos com múltiplas durações, em que os indivíduos observados, desde o início da observação até ao seu final, percorrem diversos estados. Os modelos que serão abordados neste estudo têm as seguintes características (a) todos os indivíduos iniciam o processo no mesmo estado, apesar de o fazerem em momentos diferentes, (b) para a saída de cada estado apenas existe uma transição possível e (c) a sequência dos vários estados é comum para todos os indivíduos.

O capítulo está organizado da seguinte forma:

Na secção 5.1 iremos derivar a aproximação linear da função densidade para múltiplas durações, que será o ponto de partida para, na secção 5.2, se construir uma estatística de teste para detecção da heterogeneidade negligenciada.

Na secção 5.3 serão apresentadas aproximações para a inconsistência do estimador GMM quando a heterogeneidade é negligenciada na estimação, sendo analisados alguns modelos como exemplo.

Um estimador GMM corrigido do enviesamento na condição de momentos será proposto na secção 5.4.

### 5.1 Aproximações lineares

Consideremos um processo de estados múltiplos com destino único, descrito através do tempo de permanência de cada indivíduo em cada um dos estados. Seja a sequência do tempo de permanência em cada estado dado por  $\mathbf{T} = \{T_k\}_{k=1}^R$ . Assu-

mando que é conhecida a distribuição do processo sem heterogeneidade, a sua função densidade conjunta,  $f(t_1, t_2, \dots, t_R)$ , pode ser escrita como o produto das densidades condicionais:

$$f(t_1, t_2, \dots, t_R) = f(t_1)f(t_2|t_1)\dots f(t_{R-1}|t_{R-2}, \dots, t_2, t_1)f(t_R|t_{R-1}, \dots, t_2, t_1) \quad (5.1)$$

Consideremos  $U_k$  a distribuição da heterogeneidade de cada um dos  $R$  estados. Reescrevendo  $U_k = e^{V_k}$ , com  $\mathbf{V} = \{V_k\}_{k=1}^R$ , vamos considerar que, tal como para o caso de um único estado, a variável da heterogeneidade faz parte do parâmetro de escala da distribuição da duração satisfazendo as seguintes características:

$$E(V_k) = 0, \quad \text{Var}(V_k) = \sigma_k^2, \quad E(V_k V_l) = \rho_{kl} \quad k = l = 1, 2, \dots, R$$

Assumindo que o vector  $\mathbf{U}$  de heterogeneidade negligenciada é independente da duração e dos regressores, uma aproximação à densidade conjunta dos  $R$  estados onde é incluída a heterogeneidade, pode ser deduzida por indução através das aproximações para dois, três e quatro estados.

Sendo  $f(\mathbf{t}) = f(t_1^{\alpha_1}, t_2^{\alpha_2}, \dots, t_R^{\alpha_R})$  e  $\sigma_{kl} = \sigma_k \sigma_l \rho_{kl}$ , onde  $\alpha_k$  é o *shape parameter* da cada um dos  $R$  estados, a aproximação para a função densidade conjunta com heterogeneidade é dada por

$$\begin{aligned} f_a(\mathbf{t}) = & f(\mathbf{t}) + \frac{1}{2} \sum_{k=1}^R \frac{\sigma_k^2}{\alpha_k^2} (f(\mathbf{t}) + 3t_k f^{(k)}(\mathbf{t}) + t_k^2 f^{(kk)}(\mathbf{t})) + \\ & + \sum_{k=1}^{R-1} \sum_{l=k+1}^R \frac{\sigma_{kl}}{\alpha_k \alpha_l} (f(\mathbf{t}) + t_k t_l f^{(kl)}(\mathbf{t})) + \sum_{k=1}^R \sum_{l \neq k}^R \frac{\sigma_{kl}}{\alpha_k \alpha_l} t_k f^{(k)}(\mathbf{t}) \quad (5.2) \end{aligned}$$

onde  $f^{(k)}(\mathbf{t}) = \frac{df(\mathbf{t})}{dt_k}$  e  $f^{(kl)}(\mathbf{t}) = \frac{d^2 f(\mathbf{t})}{dt_k dt_l}$ . Esta expressão é semelhante à obtida em Dumangane (2001) com  $\sigma_k^2$  e  $\sigma_{kl}$  substituídos respectivamente por  $\frac{\sigma_k^2}{\alpha_k^2}$  e  $\frac{\sigma_{kl}}{\alpha_k \alpha_l}$ .



## 5.2 Teste à heterogeneidade negligenciada

Como efectuado por Dumangane (2001), podemos derivar um teste à heterogeneidade negligenciada para o modelo com múltiplas durações e destino único. Neste caso, é um teste à hipótese conjunta de  $H_0 : Q\sigma = 0$ , onde  $\sigma$  é o vector dos  $((R + R(R - 1)/2) \times 1)$  elementos distintos da matriz  $(R \times R)$  das variâncias e covariâncias da heterogeneidade negligenciada,  $\Sigma$ , e  $Q$  é uma matriz  $(R^* \times (R + R(R - 1)/2))$  de restrições lineares que selecciona quais os  $R^*$  elementos do vector  $\sigma$  que se pretende testar.

A matriz de covariâncias,  $\Sigma$ , é composta por dois tipos distintos de elementos: os  $R$  elementos da diagonal principal,  $\sigma_k^2$ , que representam a variância da heterogeneidade em cada estado e os  $R(R - 1)/2$  elementos fora da diagonal principal,  $\sigma_{kl}$ , com  $k \neq l$ , que representam as covariâncias entre as heterogeneidades de diferentes estados.

Tal como efectuado no capítulo anterior para um estado, vamos utilizar a aproximação da função densidade para múltiplos estados dada por (5.2) como base para a construção da função verosimilhança aproximada para os parâmetros dos vectores  $\phi$  e  $\sigma$ , que é obtida através do logaritmo da aproximação da função densidade. O vector das derivadas parciais da função verosimilhança aproximada em ordem a  $\sigma$ , avaliadas em  $\sigma = 0$ , irá constituir a base para a construção da estatística de teste para detecção da heterogeneidade negligenciada.

Esse vector, denominado por  $D_\sigma(t)$ , vai ter dois tipos distintos de contribuições: a contribuição associada às  $R$  variâncias da heterogeneidade,

$$D_{\sigma_k^2}(\mathbf{t}) = \frac{1}{\alpha_k^2} \left( 1 + 3t_k \tilde{f}^{(k)}(\mathbf{t}) + t_k^2 \left( \tilde{f}^{(kk)}(\mathbf{t}) + \tilde{f}^{(k)}(\mathbf{t})^2 \right) \right), \quad k = 1, 2, \dots, R$$

e a contribuição associada às  $R(R-1)/2$  covariâncias entre as heterogeneidades de diferentes estados,

$$D_{\sigma_{kl}}(\mathbf{t}) = \frac{1}{\alpha_k \alpha_l} \left( 1 + t_k \tilde{f}^{(k)}(\mathbf{t}) + t_l \tilde{f}^{(l)}(\mathbf{t}) + t_k t_l \left( \tilde{f}^{(kl)}(\mathbf{t}) + \tilde{f}^{(k)}(\mathbf{t}) \tilde{f}^{(l)}(\mathbf{t}) \right) \right)$$

com  $k = 1, 2, \dots, R$  e  $l = 1, 2, \dots, R$ .

O vector score sob a hipótese nula é dado por

$$D_{\theta}(\mathbf{t}) = \begin{bmatrix} D_{\phi}(\mathbf{t}) \\ D_{\sigma}(\mathbf{t}) \end{bmatrix}$$

sendo a sua variância assintótica,  $\Omega$ , dada por

$$\Omega = E[D_{\theta}(\mathbf{t})D_{\theta}(\mathbf{t})']$$

Como efectuado para o modelo com um só estado, vamos definir

$$\hat{D}_{\theta}^n = n^{-1/2} \begin{bmatrix} \sum_{i=1}^n \hat{D}_{\phi}(\mathbf{t}_i) \\ \sum_{i=1}^n \hat{D}_{\sigma}(\mathbf{t}_i) \end{bmatrix}$$

onde  $\hat{D}_{\theta}(\mathbf{t})$  é igual a  $D_{\theta}(\mathbf{t})$  com os parâmetros desconhecidos substituídos pelos seus estimadores de Máxima Verosimilhança, obtidos sob a hipótese nula  $H_0 : \sigma = 0$ .

A estatística do teste score para  $H_0 : Q\sigma = 0$  é dada por

$$T^{(R^*)} = \hat{D}_{\theta}^{n'} B' (B \hat{\Omega} B')^{-1} B \hat{D}_{\theta}^n \sim \chi_{(R^*)}^2$$

sendo

$$B = \begin{bmatrix} I_p & 0 \\ 0 & Q \end{bmatrix}$$

onde  $I_p$  é uma matriz identidade ( $p \times p$ ) e  $\hat{\Omega}$  é o estimador de  $\Omega$ .

Como exemplo, vamos considerar um modelo com dois estados.  $T_1$  e  $T_2$  re-

presentam a duração completa do primeiro e segundo estados,  $\sigma_1^2$  e  $\sigma_2^2$  a variância da heterogeneidade do primeiro e segundo estados respectivamente e  $\sigma_{12}$  a covariância entre a heterogeneidade dos dois estados.

Neste modelo, o teste à heterogeneidade negligenciada é um teste à hipótese nula  $H_0 : Q\sigma = 0$ , sendo  $Q$  a matriz  $(R^* \times 3)$  das  $R^*$  restrições que se pretendem testar e  $\sigma$  o vector das covariâncias  $\sigma = [\sigma_1^2 \quad \sigma_{12} \quad \sigma_2^2]'$ .

Se pretendermos efectuar um teste à hipótese da existência de heterogeneidade negligenciada nos vários estados,  $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_{12} = 0$ , este será um teste  $T^{(3)}$ , sendo a matriz  $Q$  uma matriz identidade  $(3 \times 3)$ . A estatística de teste irá seguir uma distribuição  $\chi_{(3)}^2$ .

Por outro lado, se pretendermos efectuar um teste à hipótese de existência de heterogeneidade apenas no primeiro estado, será um teste  $T^{(2)}$  à hipótese nula  $H_0 : \sigma_1^2 = \sigma_{12} = 0$ , pelo que a matriz  $Q$  será a matriz  $(2 \times 3)$

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

e a estatística de teste irá seguir uma distribuição  $\chi_{(2)}^2$ .

### 5.3 Enviesamento dos estimadores

Consideremos que a verdadeira distribuição da duração dos  $R$  estados do modelo é gerada pela densidade conjunta  $f(\mathbf{t}, \sigma)$ . Seja  $f(\mathbf{t}) = f(t_1^{\alpha_1}, t_2^{\alpha_2}, \dots, t_R^{\alpha_R})$ , e  $\sigma$  o vector dos  $((R + R(R-1)/2) \times 1)$  elementos distintos da matriz das variâncias e covariâncias da heterogeneidade, onde  $\sigma_k^2$  corresponde aos elementos da diagonal principal e  $\sigma_{kl}$  com  $k = l = 1, 2, \dots, R$ , aos elementos fora da diagonal principal. O vector dos verdadeiros parâmetros deste modelo é dado por  $\theta_0 = \{\phi_0, \sigma_0\}$ , sendo  $\sigma_{0,k}^2$  e  $\sigma_{0,kl}$  os

elementos de  $\sigma_0$ .

Seja a condição de momentos do modelo em que é negligenciada a heterogeneidade dada por

$$\mathbf{g}(\mathbf{t}, \phi) : \quad E_{\phi}[\mathbf{g}(\mathbf{T}, \phi_0)] = 0$$

onde a sequência das durações nos vários estados é dada por  $\mathbf{T} = \{T_k\}_{k=1}^R$ .

Tal como no modelo para uma duração, a verdadeira condição de momentos é dada por  $E_{\theta_0}[\mathbf{g}(\mathbf{T}, \tilde{\phi}(\theta_0))] = 0$ , sendo a aproximação do limite em probabilidade do estimador dos verdadeiros parâmetros,  $\tilde{\phi}(\theta_0)$ , obtida por aproximação em série de Taylor em torno de  $\sigma = 0$ . Considerando que  $\phi_0 = \tilde{\phi}(\phi_0, 0)$ , essa aproximação é dada por

$$\tilde{\phi}(\theta_0) \simeq \phi_0 + \sum_{k=1}^R \left. \frac{d\tilde{\phi}(\theta_0)}{d\sigma_k^2} \right|_{\sigma=0} \sigma_{0,k}^2 + \sum_{k=1}^{R-1} \sum_{l=k+1}^R \left. \frac{d\tilde{\phi}(\theta_0)}{d\sigma_{kl}^2} \right|_{\sigma=0} \sigma_{0,kl} \quad (5.3)$$

Utilizando a aproximação conjunta para estados múltiplos dada por (5.2) e tal como foi efectuado para o modelo com um estado, vamos diferenciar a equação (4.8) em ordem a  $\tilde{\phi}(\theta_0)$  e a  $\sigma_0$  avaliada em  $\sigma = 0$ . Resolvendo as equações obtidas por aplicação do teorema da função implícita e substituindo em (5.3) obtemos a aproximação do limite em probabilidade do estimador GMM que ignora a heterogeneidade, que é dada por

$$\tilde{\phi}(\theta_0) \simeq \phi_0 - \sigma_0^2 G(\phi_0)^{-1} b(\phi_0, \sigma) \quad (5.4)$$

onde

$$G(\phi_0) = \int_0^\infty \dots \int_0^\infty \nabla_\phi \mathbf{g}(\mathbf{t}, \phi_0) f(\mathbf{t}, \phi_0) dt_R \dots dt_1$$

$$b(\phi_0, \sigma) = \sum_{k=1}^R \sigma_k^2 b^k(\phi_0) + \sum_{k=1}^{R-1} \sum_{l=k+1}^R \sigma_{kl} b^{kl}(\phi_0)$$

sendo

$$b^k(\phi_0) = \int_0^\infty \dots \int_0^\infty g(\mathbf{t}, \phi_0) \nabla_{\sigma_k^2} f_a(\mathbf{t}, \phi_0) dt_R \dots dt_1 \quad (5.5)$$

$$b^{kl}(\phi_0) = \int_0^\infty \dots \int_0^\infty g(\mathbf{t}, \phi_0) \nabla_{\sigma_{kl}} f_a(\mathbf{t}, \phi_0) dt_R \dots dt_1 \quad (5.6)$$

### 5.3.1 Exemplos

Tal como foi efectuado para os modelos com uma duração, vamos apresentar alguns exemplos para mostrar a importância que as aproximações assintóticas podem ter na análise do efeito da heterogeneidade negligenciada.

Nos exemplos aqui apresentados, as durações seguem uma distribuição exponencial, sendo referentes a modelos com duas durações ( $R = 2$ ).

#### Exemplo 1: processo de Poisson

Este é o caso mais simples de estados múltiplos em que as durações dos vários estados são independentes e igualmente distribuídas através de uma distribuição exponencial de parâmetro escala  $\lambda_k$  tal que

$$\log \lambda_k = \beta_0 + \beta_1' x, \quad k = 1, 2$$

Consideremos o vector de parâmetros  $\beta = \{\beta_0, \beta_1\}$ , onde o primeiro componente é um escalar e o segundo é um vector de parâmetros associados ao conjunto de regressores  $x$ . Tal como anteriormente,  $\beta_0$  é definido de forma que  $x$  tenha média nula e matriz de covariâncias  $\Sigma$ .

Através da aplicação de (5.4) obtemos a aproximação do limite em probabilidade do estimador de Máxima Verosimilhança

$$\begin{bmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{bmatrix} \simeq \begin{bmatrix} \beta_0 - \frac{1}{4}(\sigma_1^2 + \sigma_2^2) \\ \beta_1 \end{bmatrix}$$

Pode-se verificar que, apesar de ignorar a presença da heterogeneidade nas duas durações, a estimação por Máxima Verosimilhança produz estimadores consistentes dos coeficientes dos regressores. No entanto, a estimação do termo independente é enviesada para um valor inferior ao do verdadeiro parâmetro.

## **Exemplo 2: *Ocurrence Dependence***

Vamos agora considerar uma extensão ao modelo apresentado no exemplo anterior. Neste modelo, as durações dos vários estados continuam a ser independentes mas os termos independentes e os coeficientes dos regressores variam entre os vários estados. Este modelo é definido por

$$\log \lambda_k = \beta_{0k} + \beta'_{1k}x, \quad k = 1, 2$$

Assumindo as mesmas condições de distribuição dos regressores e dos termos independentes do exemplo anterior, a aproximação do limite em probabilidade dos

estimadores de Máxima Verosimilhança é dado por

$$\begin{bmatrix} \tilde{\beta}_{01} \\ \tilde{\beta}_{11} \\ \tilde{\beta}_{02} \\ \tilde{\beta}_{12} \end{bmatrix} \simeq \begin{bmatrix} \beta_{01} - \frac{1}{2}\sigma_1^2 \\ \beta_{11} \\ \beta_{02} - \frac{1}{2}\sigma_2^2 \\ \beta_{12} \end{bmatrix}$$

Tal como no processo de Poisson, também neste caso a estimação por Máxima Verosimilhança, apesar de ter sido ignorada a presença de heterogeneidade, produz estimadores consistentes dos coeficientes dos regressores, sendo a estimação dos termos independentes de ambos os estados enviesados para um valor inferior ao verdadeiro valor.

### Exemplo 3: dependência temporal

Consideremos um modelo em que a função *hazard*, condicional em  $x$ , de um estado depende da duração dos estados anteriores. Este modelo, para dois estados, é definido por

$$\log \lambda_1 = \beta_{01} + \beta'_{11}x$$

$$\log \lambda_2 = \beta_{02} + \beta'_{12}x + \gamma \log t_1$$

onde o parâmetro  $\gamma$ , o coeficiente da dependência da duração, faz com que esses estados sejam dependentes.

O limite em probabilidade dos estimadores de Máxima Verosimilhança para os

parâmetros  $\gamma$  e  $\beta_k = \{\beta_{0k}, \beta_{1k}\}$  com  $k = 1, 2$  é dado por

$$\begin{bmatrix} \tilde{\gamma} \\ \tilde{\beta}_{01} \\ \tilde{\beta}_{11} \\ \tilde{\beta}_{02} \\ \tilde{\beta}_{12} \end{bmatrix} \simeq \begin{bmatrix} \gamma - \frac{\sigma_{12}}{\psi'(1)} \\ \beta_{01} - \frac{\sigma_1^2}{2} \\ \beta_{11} \\ \beta_{02} - \frac{\sigma_2^2}{2} - \frac{\beta_{01} - \psi(1)}{\psi'(1)} \sigma_{12} \\ \beta_{12} - \beta_{11} \frac{\sigma_{12}}{2} \end{bmatrix}$$

Devido ao primeiro estado não estar dependente de nenhum outro, verifica-se que o enviesamento dos seus parâmetros é semelhante ao enviesamento verificado no exemplo anterior. O enviesamento dos restantes parâmetros, referentes ao segundo estado, pode assumir diversas formas:

- O enviesamento do parâmetro da dependência da duração,  $\gamma$ , pode ser para valores inferiores ou superiores ao verdadeiro valor, dependendo se a correlação entre as distribuições de heterogeneidade é, respectivamente, positiva ou negativa.
- O coeficiente dos regressores está dependente da correlação entre as distribuições da heterogeneidade e do coeficiente dos regressores do estado anterior, pelo que o seu enviesamento é para valores inferiores aos reais se esses dois parâmetros de que depende tiverem o mesmo sinal, tendo um enviesamento para valores superiores no caso de terem sinais opostos.
- A estimação do termo independente é muito complexa, estando dependente do valor do termo independente do primeiro estado, da variância da heterogeneidade do segundo estado e da covariância entre as distribuições de heterogeneidade.

É também fácil verificar que, caso não exista correlação entre as distribuições de heterogeneidade, a covariância entre elas é nula e o enviesamento dos coeficientes



dos regressores é semelhante ao exemplo anterior.

## 5.4 Estimador GMM corrigido

Tal como foi efectuado para os modelos com uma duração, vamos derivar um estimador GMM para os modelos com múltiplas durações.

A condição de momentos corrigida é tal que  $E_{\theta}[\mathbf{g}(\mathbf{T}, \boldsymbol{\phi}) - \sigma^2 \mathbf{b}(\mathbf{T}, \boldsymbol{\phi}, \boldsymbol{\sigma})] \simeq 0$ , onde  $\sigma^2 \mathbf{b}(\mathbf{t}, \boldsymbol{\phi})$ , a aproximação da função enviesamento estrutural que, tal como demonstrado no apêndice B, é dada por

$$\mathbf{b}(\mathbf{t}, \boldsymbol{\phi}_0, \boldsymbol{\sigma}) = \frac{1}{2} \sum_{k=1}^R \frac{\sigma_k^2}{\alpha_k^2} (t_k \mathbf{g}^k(\mathbf{t}, \boldsymbol{\phi}_0) + t_k^2 \mathbf{g}^{kk}(\mathbf{t}, \boldsymbol{\phi}_0)) + \sum_{k=1}^{R-1} \sum_{l=k+1}^R \frac{\sigma_{kl}}{\alpha_k \alpha_l} t_k t_l \mathbf{g}^{kl}(\mathbf{t}, \boldsymbol{\phi}_0)$$

onde  $\mathbf{g}^k(\mathbf{t}, \boldsymbol{\phi}_0) = \frac{d\mathbf{g}(\mathbf{t}, \boldsymbol{\phi}_0)}{dt_k}$  e  $\mathbf{g}^{kl}(\mathbf{t}, \boldsymbol{\phi}_0) = \frac{d^2\mathbf{g}(\mathbf{t}, \boldsymbol{\phi}_0)}{dt_k dt_l}$ .

Facilmente se verifica que quando  $R = 1$  a aproximação da função estrutural do enviesamento coincide com a função para uma duração dada por (4.26).

## 6. Simulações

Neste capítulo vamos proceder a simulações de Monte Carlo para analisar o comportamento do estimador GMM corrigido na presença de heterogeneidade.

Para elaboração das simulações, foram considerados três tipos distintos de modelos:

1. Modelo com observações completas e sem regressores.
2. Modelo com observações completas e com um regressor.
3. Modelo com observações censuradas baseadas num mecanismo de censura aleatória e independente e com um regressor.

Em todos os modelos, a heterogeneidade negligenciada é dada por  $u = e^v$ , com  $E(V) = 0$  e com  $V$  distribuído independente de  $T$ , seguindo duas distribuições:  $\text{Gamma}(\alpha, \beta)$  e  $\text{Lognormal}(\mu, \sigma_{LN}^2)$ . A escolha da distribuição Gamma para a heterogeneidade deve-se ao facto desta ser a distribuição preferida por muitos autores e softwares estatísticos (ver Abbrigo e Van den Berg, 2003). A opção pela distribuição Lognormal deve-se ao facto da heterogeneidade, sendo a soma de diversos factores não observáveis, poder ter uma distribuição Normal.

Os parâmetros das distribuições foram escolhidos de forma a que

$$\rho = \frac{\text{Var}(\log T)}{\text{Var}(\log TH)} \in \{0.8, 0.9\}$$

No caso das distribuições Weibull e Log-logistic, o valor de  $\rho$  é dado respectivamente por

$$\rho_W = \frac{\psi'(1)}{\psi'(1) + \sigma_v^2} \quad \rho_{LL} = \frac{2\psi'(1)}{2\psi'(1) + \sigma_v^2}$$

A tabela 6.1 mostra os valores dos parâmetros para cada uma das distribuições conforme o valor escolhido para  $\rho$ . Na última coluna é apresentada a aproximação do limite em probabilidade para o *shape parameter* calculado com base em (4.21) para o modelo Weibull e em (4.24) para o modelo Log-logistic.

Tabela 6.1: Parâmetros das distribuições de heterogeneidade

Distribuição	$\rho$	$\sigma_v^2$	$\sigma_{LN}^2$	$\sigma_G^2$	$\tilde{\alpha}/\alpha$
Weibull	0.8	0.411	0.414	0.345	0.800
	0.9	0.183	0.199	0.168	0.900
Log-logistic	0.8	0.822	0.843	0.605	0.874
	0.9	0.366	0.368	0.312	0.940

Para cada um dos modelos foram utilizadas duas distribuições de  $T$ : a distribuição Weibull com a função sobrevivência

$$S(t, \alpha, \lambda) = \exp(-\lambda t^\alpha)$$

e a distribuição Log-logistic com a função sobrevivência

$$S(t, \alpha, \lambda) = (1 + \lambda t^\alpha)^{-1}$$

sendo o parâmetro de escala, em ambos os casos, dado por  $\lambda = \exp(x'\beta)$ .

As simulações foram efectuadas com 2000 réplicas de amostras com 200 e 500 observações.

Foram efectuadas simulações para a estimação por Máxima Verosimilhança e pelo estimador GMM corrigido. Neste último caso, foram considerados dois tipos de estimadores: o primeiro, GMMR (Generalised Method of Moments Restricted), assume que o valor da variância da heterogeneidade é conhecida no momento da estimação e utiliza como condição de momentos as funções score associadas à densidade

de  $T$  corrigidas do enviesamento; o segundo estimador, GMMU (Generalised Method of Moments Unrestricted), utiliza a função score da variância da heterogeneidade avaliada em  $\sigma^2 = 0$  e corrigida do enviesamento como uma condição de momentos adicional para a estimação da variância da heterogeneidade.

As funções aproximadas das condições de momentos corrigidas para o estimador GMMR foram derivadas tal como referido nas secções anteriores, tendo-se obtido as seguintes funções para a distribuição Weibull

$$\begin{aligned} g_{\alpha}^c(t, x, \theta) &= \frac{1 + \sigma^2 z(t)}{\alpha} + \left(1 - z(t) \exp\left(-\frac{\sigma^2}{2}\right)\right) \log(t) \\ g_{\beta}^c(t, x, \theta) &= \left(1 - z(t) \exp\left(-\frac{\sigma^2}{2}\right)\right) x \end{aligned} \quad (6.1)$$

e para a distribuição Log-logistic

$$\begin{aligned} g_{\alpha}^c(t, x, \theta) &= \left(\frac{1}{\alpha} + \frac{(1 - z(t)) \log(t)}{1 + z(t)}\right) \exp\left(\frac{z(t)\sigma^2}{(1 + z(t))^2}\right) + \frac{z(t)\sigma^2}{(1 + z(t))^2 \alpha} \\ g_{\beta}^c(t, x, \theta) &= \left(\frac{1 - z(t)}{1 + z(t)} \exp\left(\frac{z(t)\sigma^2}{(1 + z(t))^2}\right)\right) x \end{aligned} \quad (6.2)$$

onde  $z(t) = t^{\alpha} \exp(x'\beta)$  e  $\theta$  é o vector de parâmetros a estimar.

Para a estimação por GMMU, tivemos de derivar a aproximação da condição de momentos corrigida para a variância da heterogeneidade. Para a distribuição Weibull essa aproximação é dada por

$$g_{\sigma^2}^c(t, x, \theta) = 1 - 3z(t) \exp\left(-\frac{\sigma^2}{2}\right) + z(t)^2 \exp(-2\sigma^2) \quad (6.3)$$

e para a distribuição Log-logistic por

$$g_{\sigma^2}^c(t, x, \theta) = \frac{1 - 4z(t) + z(t)^2}{(1 + z(t))^2} \exp\left(\frac{3z(t)\sigma^2}{(1 + z(t))^2}\right) \quad (6.4)$$

## 6.1 Modelo com observações completas sem regressores

Neste modelo, o parâmetro de escala  $\lambda = \exp(x'\beta)$  é reparameterizado de forma a que  $\beta = 0$  e  $x = 1$ .

Na simulação com o estimador de Máxima Verosimilhança, os resultados obtidos para o termo independente e para o *shape parameter* não sofreram alterações para diferentes valores de  $\alpha$ . Este resultado está de acordo com o obtido em (4.21) e (4.24), onde se pode ver que o enviesamento do *shape parameter* apenas depende do valor de  $\sigma^2$ . Da mesma forma, na simulação com o estimador GMMR, onde se assume que  $\sigma^2$  é conhecido, os resultados não sofreram alterações para diferentes valores de  $\alpha$ . Assim, os resultados apresentados para os estimadores de Máxima Verosimilhança e GMMR são válidos para qualquer valor de  $\alpha$ .

No entanto, na simulação com o estimador GMMU, como  $\sigma^2$  não é conhecido, a sua estimação vai estar dependente do valor estimado para o *shape parameter*, pelo que, ao contrário dos estimadores de Máxima Verosimilhança e GMMR, os resultados serão diferentes para diferentes valores de  $\alpha$ , de forma que serão analisados os resultados para  $\alpha \in \{0.8, 1, 1.5\}$  quando  $\rho = 0.8$  e  $\alpha = 1$  quando  $\rho = 0.9$ .

Na tabela 6.2 podemos ver o enviesamento do estimador de Máxima Verosimilhança quando  $\rho \in \{0.8, 0.9\}$ . O enviesamento está medido em valor absoluto para o termo independente,  $\beta_0$ , e proporcional para o *shape parameter*,  $\alpha$ .

Através da análise da tabela, podemos verificar que:

- O enviesamento dos parâmetros é sempre para valores inferiores ao valor do verdadeiro parâmetro, quer no termo independente, quer no *shape parameter*, sendo esse enviesamento claramente superior quando utilizamos o modelo Weibull.

Tabela 6.2: Resultados da simulação dos modelos Weibull e Log-logistic sem regressores: estimador de Máxima verosimilhança

			Máxima Verosimilhança							
			Weibull				Log-logistic			
			$\hat{\beta}_0 - \beta_0$		$\hat{\alpha} / \alpha$		$\hat{\beta}_0 - \beta_0$		$\hat{\alpha} / \alpha$	
$\rho$	n	U	média	dp	média	dp	média	dp	média	dp
0.8	200	LN	-.076	.075	.830	.048	.000	.130	.887	.051
		G	-.076	.076	.808	.055	.014	.128	.889	.050
	500	LN	-.074	.047	.825	.030	.005	.080	.884	.032
		G	-.075	.048	.802	.036	.016	.079	.886	.032
0.9	200	LN	-.039	.077	.914	.053	-.002	.129	.948	.055
		G	-.039	.078	.908	.054	.001	.128	.949	.056
	500	LN	-.038	.046	.909	.033	-.002	.080	.945	.034
		G	-.038	.047	.901	.033	.002	.079	.945	.035

- Através da comparação com os resultados obtidos na tabela 6.1 para a aproximação do limite em probabilidade do *shape parameter*, verifica-se que a aproximação dá uma boa previsão do seu enviesamento, excepto para o modelo Weibull com a distribuição da heterogeneidade Lognormal e com  $\rho = 0.8$ .
- O enviesamento do termo independente é quase nulo no modelo Log-logistic, o mesmo já não se verificando no modelo Weibull. Estes resultados estão de acordo com o enviesamento da condição de momentos constatado nas figuras 4.1 e 4.2.
- O enviesamento do *shape parameter* no modelo Log-logistic não parece ser muito sensível à distribuição da heterogeneidade visto não apresentar diferenças significativas quando se utilizam diferentes distribuições de heterogeneidade, ao contrário dos resultados obtidos no modelo Weibull, que parece ser sensível à distribuição da heterogeneidade.

A tabela 6.3 apresenta o resultado das simulações para os modelos Weibull e Log-logistic com o estimador GMM corrigido da heterogeneidade, quando  $\sigma^2$  é conhecido.

Tabela 6.3: Resultados da simulação dos modelos Weibull e Log-logistic sem regressores: estimador GMMR

			GMMR							
			Weibull				Log-logistic			
			$\beta_0 - \beta_0$		$\hat{\alpha}/\alpha$		$\beta_0 - \beta_0$		$\hat{\alpha}/\alpha$	
$\rho$	n	U	média	dp	média	dp	média	dp	média	dp
0.8	200	LN	-.036	.095	1.050	.066	.000	.147	1.002	.058
		G	-.033	.095	1.014	.078	.020	.146	1.005	.058
	500	LN	-.032	.060	1.043	.042	.005	.091	1.000	.036
		G	-.029	.060	1.003	.051	.023	.090	1.002	.036
0.9	200	LN	-.006	.086	1.018	.060	-.002	.137	1.006	.058
		G	-.005	.087	1.010	.062	.002	.136	1.007	.060
	500	LN	-.005	.052	1.012	.037	-.002	.085	1.002	.036
		G	-.005	.052	1.002	.038	.002	.084	1.003	.037

Através da sua análise e, comparando com o estimador de Máxima Verosimilhança, verifica-se que:

- O enviesamento dos parâmetros diminui nos dois modelos, apesar da correcção ser melhor no modelo Log-logistic, em que a média proporcional do *shape parameter* é bastante próxima de 1, estando de acordo com o constatado nas figuras 4.1 e 4.2.
- No modelo Log-logistic, excepto para o termo independente e quando  $\rho = 0.8$ , continua a não se verificarem diferenças significativas entre os resultados obtidos utilizando diferentes distribuições de heterogeneidade. No modelo Weibull essa diferença mantém-se, sendo o modelo com distribuição de heterogeneidade Gamma a produzir melhores resultados para o *shape parameter*.

A tabela 6.4 mostra os resultados obtidos para a simulação com o modelo Weibull utilizando o estimador GMM corrigido com  $\sigma^2$  desconhecido.

Através da análise dos resultados apresentados na tabela, pode-se verificar que:

- O enviesamento do termo independente é afectado pelo valor de  $\rho$  e de  $\alpha$  de forma inversa, aumentando o enviesamento à medida que se verifica a sua diminuição.
- O enviesamento do *shape parameter* é bastante sensível ao seu verdadeiro valor.

Tabela 6.4: Resultados da simulação do modelo Weibull sem regressores: estimador GMMU

				GMMU					
				Weibull					
				$\hat{\beta}_0 - \beta_0$		$\hat{\alpha}/\alpha$		$\hat{\sigma}^2/\sigma^2$	
$\alpha$	$\rho$	n	U	média	dp	média	dp	média	dp
1	0.8	200	LN	-.051	.086	.961	.106	.671	.416
			G	-.058	.086	.932	.140	.754	.336
		500	LN	-.043	.055	.969	.084	.744	.321
			G	-.053	.055	.909	.116	.770	.280
	0.9	200	LN	-.014	.086	1.011	.103	.981	.887
			G	-.012	.086	1.009	.107	1.079	.859
		500	LN	-.008	.054	1.018	.084	1.090	.778
			G	-.005	.054	1.024	.087	1.281	.758
0.8	0.8	200	LN	-.068	.102	1.006	.163	.821	.610
			G	-.079	.099	.962	.169	.800	.551
		500	LN	-.061	.073	1.025	.153	.933	.543
			G	-.073	.062	.927	.135	.723	.423
1.5	0.8	200	LN	-.041	.086	.955	.104	.666	.408
			G	-.037	.084	.918	.132	.702	.394
		500	LN	-.032	.055	.964	.084	.746	.317
			G	-.027	.052	.898	.116	.696	.320

- No *shape parameter*, ao contrário dos resultados obtidos com o estimador GMMR, é com a distribuição da heterogeneidade Lognormal que se obtém melhores resultados.

Relativamente aos resultados obtidos para a estimação da variância da heterogeneidade, pode-se constatar que:

- O seu valor é sempre subestimado, excepto quando  $\rho = 0.9$ .
- A precisão da estimação deste parâmetro aumenta à medida que  $\rho$  diminui. No entanto, quando  $\alpha = 0.8$ , verifica-se uma perda de precisão.

Na tabela 6.5 são apresentados os resultado para o modelo Log-logistic quando é utilizado o estimador GMMU corrigido com  $\sigma^2$  desconhecido.

A sua análise permite-nos verificar que:

- Tal como verificado para os outros estimadores, as características da distribuição da heterogeneidade parecem não ter efeito na estimação do *shape parameter*.



Tabela 6.5: Resultados da simulação do modelo Log-logistic sem regressores: estimador GMMU

				GMMU					
				Log-logistic					
				$\hat{\beta}_0 - \beta_0$		$\hat{\alpha} / \alpha$		$\hat{\sigma}^2 / \sigma^2$	
$\alpha$	$\rho$	n	U	média	dp	média	dp	média	dp
1	0.8	200	LN	.001	.138	.943	.069	.495	.409
			G	.017	.136	.943	.079	.466	.495
		500	LN	.005	.085	.937	.045	.453	.286
			G	.020	.086	.953	.061	.578	.427
	0.9	200	LN	-.002	.136	.997	.073	.882	.868
			G	.002	.132	.981	.064	.571	.500
		500	LN	-.002	.083	.983	.049	.696	.639
			G	.002	.082	.989	.043	.764	.416
0.8	0.8	200	LN	.004	.135	.938	.078	.446	.492
			G	.020	.133	.943	.080	.464	.517
		500	LN	.003	.085	.946	.061	.531	.414
			G	.019	.086	.953	.062	.573	.438
1.5	0.8	200	LN	.000	.137	.934	.073	.419	.428
			G	.017	.136	.940	.076	.445	.453
		500	LN	.005	.085	.945	.058	.524	.370
			G	.020	.085	.950	.058	.557	.381

- O enviesamento do *shape parameter*, ao contrário do verificado com o modelo Weibull, é pouco sensível a variações no seu verdadeiro valor, o mesmo acontecendo com o termo independente.

No que se refere à estimação da variância da heterogeneidade:

- O enviesamento é superior ao verificado para o modelo Weibull, apesar de não se verificarem grandes alterações nos dois modelos ao nível da precisão.

As figuras 6.1 e 6.2 mostram-nos a estimativa da densidade do estimador do *shape parameter*,  $\alpha$ , pelos estimadores de Máxima Verosimilhança (MLE) e GMM corrigidos (GMMR e GMMU), na estimação do modelo Weibull e Log-logistic, respectivamente, quando a distribuição de heterogeneidade é Lognormal<sup>7</sup>.

<sup>7</sup> Apenas serão apresentados os gráficos quando a heterogeneidade tem uma distribuição Lognormal visto que, na sua maioria, são semelhantes aos gráficos obtidos quando a heterogeneidade tem uma distribuição Gamma.

Figura 6.1: Densidades dos estimadores MLE, GMMR e GMMU da estimação de  $\alpha$  no modelo Weibull com distribuição de heterogeneidade Lognormal e  $n = 500$

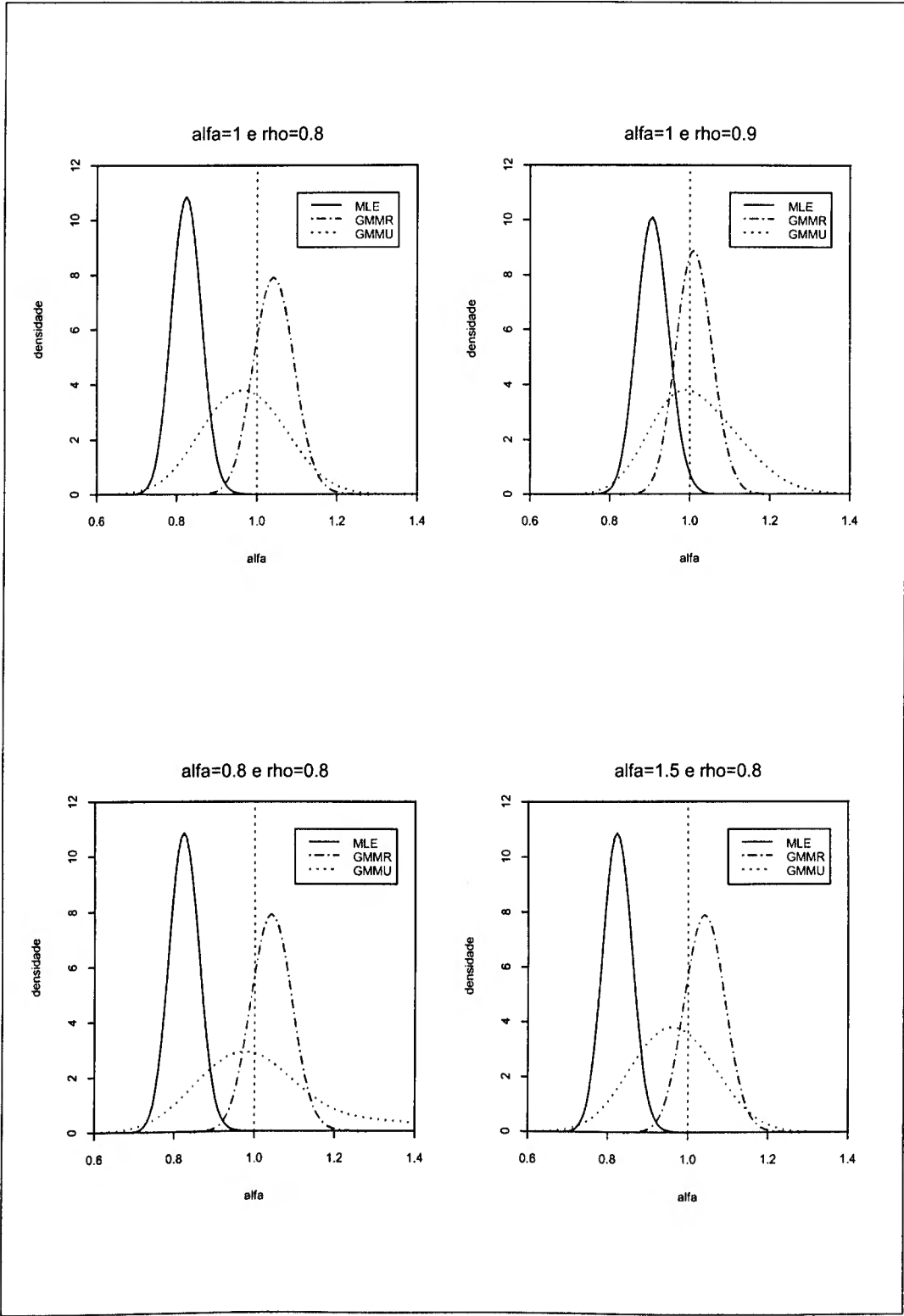
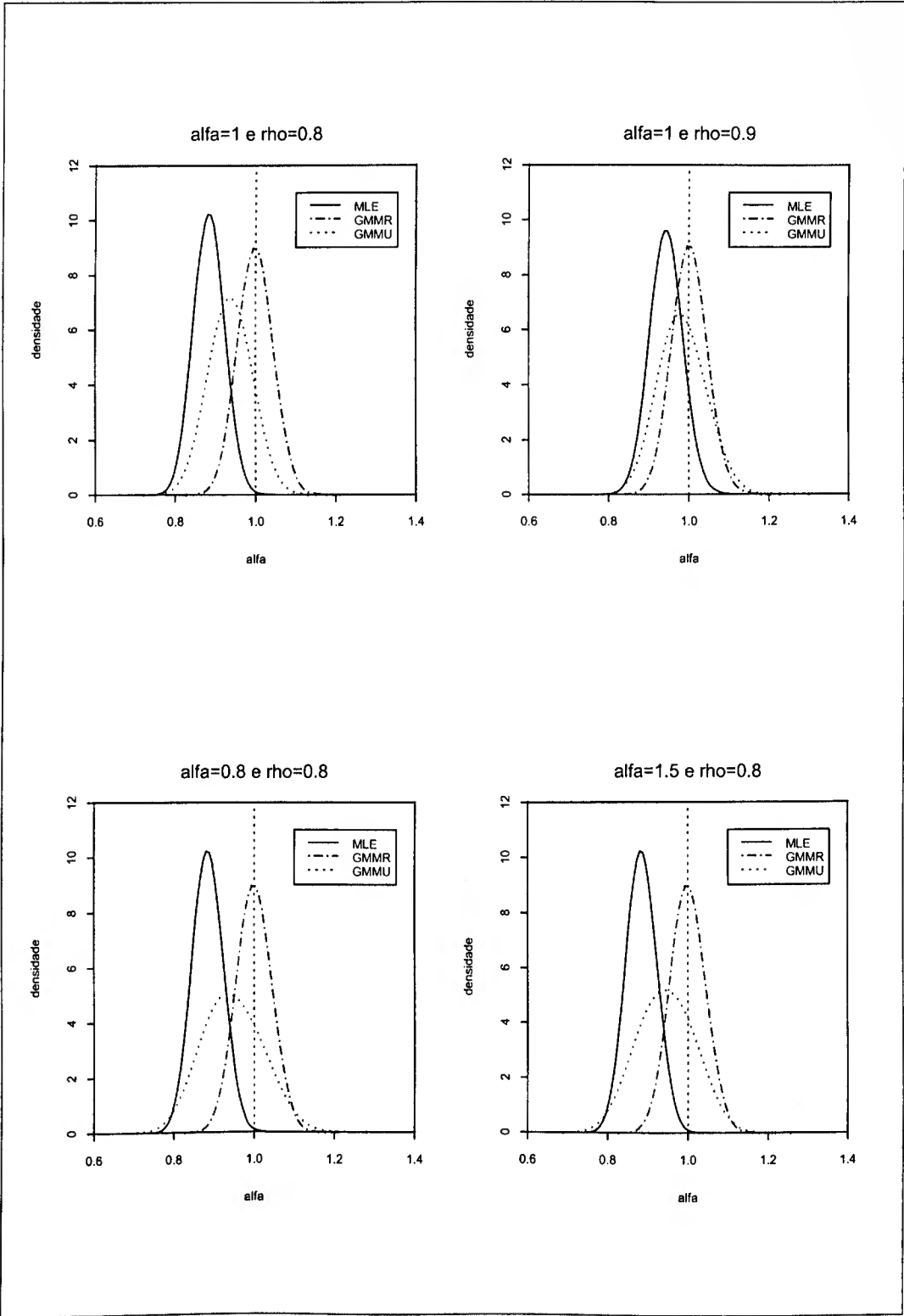


Figura 6.2: Densidades dos estimadores MLE, GMMR e GMMU da estimação de  $\alpha$  no modelo Log-logistic com distribuição de heterogeneidade Lognormal e  $n = 500$



É fácil verificar que no modelo Weibull (figura 6.1) o estimador de Máxima Verosimilhança é subavaliado, enquanto que o estimador GMMR é um pouco sobreavaliado. Comparando estes estimadores através dos vários painéis com  $\rho = 0.8$ , pode-se ver que as estimativas das suas densidades são independentes do valor do *shape parameter*,  $\alpha$ . Para o estimador GMMU, verifica-se uma clara perda de precisão aliada a uma perda de simetria em relação aos outros estimadores.

Tal como no modelo Weibull, também se pode verificar que no modelo Log-logistic (figura 6.2) as estimativas das densidades para os estimadores de Máxima Verosimilhança e GMMR são independentes do valor do *shape parameter*. Também se pode verificar que o estimador GMMR encontra-se distribuído simetricamente em torno da unidade e que a perda de precisão neste modelo é menor que no modelo Weibull. Relativamente ao estimador GMMU, verifica-se uma perda de precisão, no entanto, essa perda de precisão, é muito inferior quando comparado com o modelo Weibull.

## 6.2 Modelo com observações completas com um regressor

Através desta simulação pretendemos analisar o enviesamento nos regressores e averiguar se as previsões efectuadas para o *shape parameter* estão correctas.

Para a sua elaboração, consideramos que os dados são gerados através de uma distribuição condicional num conjunto de regressores  $x = [1, x_1]$ , ao qual está associado o conjunto de coeficientes  $\{\beta_0, \beta_1\} = \{0, 1\}$ . A análise será efectuada para  $\alpha = 1$  e  $\rho = 0.8$ .

Para evitar introduzir um novo elemento de aleatoriedade nesta experiência

utilizamos formatos de regressores fixos (fixed regressor design). Assim, o regressor  $x_1$  é constituído por 50 valores repetidos 4 ou 10 vezes, conforme o tamanho da amostra seja de 200 ou 500 observações respectivamente. Os valores são atribuídos segundo os seguintes formatos:

- Formato F1: 50 valores igualmente espaçados no intervalo  $[-1, 1]$ ;
- Formato F2: 50 valores igualmente espaçados no intervalo  $[0, 1]$ ;
- Formato F3: 50 valores constituídos pelos quadrados dos valores do formato F2.

A tabela 6.6 mostra-nos o enviesamento proporcional do estimador de Máxima Verosimilhança para o coeficiente do regressor e para o *shape parameter*.

Tabela 6.6: Resultados da simulação dos modelos Weibull e Log-logistic com regressores para  $\rho=0.8$  e  $\alpha=1$ : estimador de Máxima Verosimilhança

			Máxima Verosimilhança							
			Weibull				Log-logistic			
			$\beta_1/\beta_1$		$\hat{\alpha}/\alpha$		$\beta_1/\beta_1$		$\hat{\alpha}/\alpha$	
Formato	n	U	média	dp	média	dp	média	dp	média	dp
F1	200	LN	.838	.144	.833	.049	.882	.216	.889	.051
		G	.815	.154	.812	.055	.891	.212	.891	.051
	500	LN	.826	.090	.826	.030	.881	.132	.885	.032
		G	.805	.095	.803	.036	.879	.136	.887	.032
F2	200	LN	.843	.275	.833	.049	.874	.421	.889	.051
		G	.819	.292	.812	.055	.890	.416	.891	.051
	500	LN	.825	.172	.826	.030	.876	.260	.885	.032
		G	.807	.181	.803	.036	.872	.267	.887	.032
F3	200	LN	.845	.266	.833	.049	.874	.405	.889	.051
		G	.824	.283	.812	.055	.887	.399	.891	.051
	500	LN	.828	.166	.826	.030	.878	.252	.885	.032
		G	.809	.174	.803	.036	.874	.259	.887	.032

A sua análise permite-nos verificar que:

- o enviesamento proporcional do *shape parameter* não depende da distribuição do regressor visto que, quer o enviesamento quer o desvio padrão são iguais para todos os formatos aqui considerados e a sua diferença em relação ao modelo sem

regressores não é significativa.

- O enviesamento do coeficiente do regressor não sofre grandes alterações ao longo dos vários formatos aqui considerados, sendo os seus valores semelhantes aos obtidos no *shape parameter*, o que está de acordo com o previsto na aproximação do limite em probabilidade.
- O desvio padrão para o coeficiente do regressor é semelhante nos formatos F2 e F3 mas, nestes formatos, é superior em cerca do dobro ao desvio padrão de F1.

Na tabela 6.7 são apresentados os resultados para a simulação do estimador GMM corrigido quando  $\sigma^2$  é conhecido.

Tabela 6.7: Resultados da simulação dos modelos Weibull e Log-logistic com regressores para  $\rho=0.8$  e  $\alpha=1$ : estimador GMMR

			GMMR							
			Weibull				Log-logistic			
			$\hat{\beta}_1/\beta_1$		$\hat{\alpha}/\alpha$		$\hat{\beta}_1/\beta_1$		$\hat{\alpha}/\alpha$	
Formato	n	U	média	dp	média	dp	média	dp	média	dp
F1	200	LN	1.063	.194	1.056	.066	.996	.247	1.005	.058
		G	1.025	.212	1.021	.077	1.007	.242	1.008	.058
	500	LN	1.044	.122	1.045	.042	.995	.151	1.001	.036
		G	1.008	.132	1.006	.051	.994	.156	1.003	.036
F2	200	LN	1.069	.371	1.056	.066	.987	.464	1.005	.058
		G	1.029	.402	1.021	.077	1.007	.475	1.008	.058
	500	LN	1.043	.233	1.045	.042	.990	.297	1.001	.036
		G	1.010	.251	1.006	.051	.986	.306	1.003	.036
F3	200	LN	1.074	.357	1.056	.066	.986	.464	1.005	.058
		G	1.039	.388	1.021	.077	1.004	.456	1.008	.058
	500	LN	1.048	.225	1.045	.042	.992	.289	1.001	.036
		G	1.015	.241	1.006	.051	.988	.296	1.003	.036

Analisando os resultados, podemos verificar que:

- Os valores encontrados para o *shape parameter* não se alteram com os formatos considerados e são bastante semelhantes aos valores obtidos na estimação sem regressores, o que permite concluir que, nestes modelos, o enviesamento do *shape*

*parameter* não depende dos regressores.

- No modelo Weibull, o enviesamento proporcional do coeficiente do regressor vem sobrestimado mas, tal como no caso do *shape parameter*, essa sobrestimação tende a diminuir à medida que aumenta o tamanho da amostra.
- No modelo Log-logistic, o enviesamento proporcional do coeficiente do regressor está próximo da unidade.
- Tal como no estimador de Máxima Verosimilhança, o enviesamento proporcional do coeficiente do regressor não sofre alterações significativas ao longo dos vários formatos, apesar de se verificar uma precisão menor nos formatos F2 e F3 em relação ao formato F1.

A tabela 6.8 mostra-nos os resultados do estimador GMM corrigido com  $\sigma^2$  desconhecido para o modelo Weibull.

Tabela 6.8: Resultados da simulação do modelo Weibull com regressores para  $\rho=0.8$  e  $\alpha=1$ : estimador GMMU

			GMMU					
			Weibull					
			$\hat{\beta}_1/\beta_1$		$\hat{\alpha}/\alpha$		$\hat{\sigma}^2/\sigma^2$	
Formato	n	U	média	dp	média	dp	média	dp
F1	200	LN	.924	.213	.963	.106	.655	.435
		G	.852	.279	.944	.129	.797	.351
	500	LN	.919	.162	.971	.083	.757	.328
		G	.781	.233	.914	.115	.828	.260
F2	200	LN	.932	.429	.972	.111	.681	.451
		G	.866	.589	.968	.155	.823	.345
	500	LN	.906	.308	.979	.082	.774	.332
		G	.768	.438	.929	.111	.843	.250
F3	200	LN	.937	.401	.971	.107	.682	.444
		G	.853	.551	.960	.121	.796	.415
	500	LN	.897	.289	.977	.082	.768	.331
		G	.747	.418	.927	.112	.834	.260

Estes resultados permitem-nos verificar que:

- Apesar dos resultados obtidos para o *shape parameter* variarem entre os formatos considerados, continuam a estar próximos dos resultados do modelo sem regressores.
- Os valores do coeficiente do regressor vêm subestimados e o seu enviesamento aumenta à medida que aumenta o tamanho da amostra.
- Os resultados da estimação da variância da heterogeneidade não são muito diferentes dos obtidos na estimação sem regressores e a sua variação ao longo dos vários formatos não é muito significativa.

Na tabela 6.9 temos os resultados do estimador GMM corrigido com  $\sigma^2$  desconhecido para o modelo Log-logistic.

Tabela 6.9: Resultados da simulação do modelo Log-logistic com regressores para  $\rho=0.8$  e  $\alpha=1$ : estimador GMMU

			GMMU					
			Log-logistic					
			$\hat{\beta}_1/\beta_1$		$\hat{\alpha}/\alpha$		$\hat{\sigma}^2/\sigma^2$	
Formato	n	U	média	dp	média	dp	média	dp
F1	200	LN	.921	.240	.941	.077	.904	.963
		G	.935	.236	.946	.080	.955	1.016
	500	LN	.930	.154	.948	.060	1.094	.808
		G	.938	.161	.955	.062	1.179	.845
F2	200	LN	.896	.455	.941	.077	.903	.948
		G	.918	.447	.946	.080	.942	.983
	500	LN	.900	.286	.948	.060	.942	.983
		G	.915	.301	.955	.062	1.177	.845
F3	200	LN	.894	.435	.941	.077	.903	.946
		G	.916	.430	.946	.080	.945	.986
	500	LN	.903	.280	.949	.060	1.100	.809
		G	.917	.292	.956	.062	1.182	.852

Através da sua análise, constatamos que:

- Os resultados obtidos para o *shape parameter* mostram-nos que o seu enviesa-



mento continua a ser independente do formato do regressor e da distribuição da heterogeneidade, continuando a ter valores semelhantes aos obtidos no modelo sem regressores.

- Os valores do coeficiente do regressor vêm subestimados, sendo o seu enviesamento inferior em F1 e muito semelhante para F2 e F3.
- O enviesamento proporcional na estimação da variância da heterogeneidade está bastante próximo de 1 mas o seu desvio padrão é muito elevado.

As figuras 6.3 e 6.4 mostram-nos as estimativas da densidade da estimação do coeficiente do regressor nos modelos Weibull e Log-logistic quando a heterogeneidade segue uma distribuição Lognormal.

Pode-se verificar que no modelo Weibull (figura 6.3) os resultados obtidos para o coeficiente dos regressores nos formatos F2 e F3 são semelhantes mas, as suas dispersões são superiores à dispersão do formato F1.

Também no modelo Log-logistic (figura 6.4) verifica-se que os formatos F2 e F3 são muito semelhantes e a sua estimação tem uma maior dispersão que o formato F1. No entanto, neste modelo, a perda de precisão dos estimadores GMM corrigidos em relação à estimação por Máxima Verosimilhança é mais pequena que no modelo Weibull.

Figura 6.3: Densidades dos estimadores MLE, GMMR e GMMU da estimação de  $\beta_1$  no modelo Weibull com distribuição de heterogeneidade Lognormal e  $n = 500$

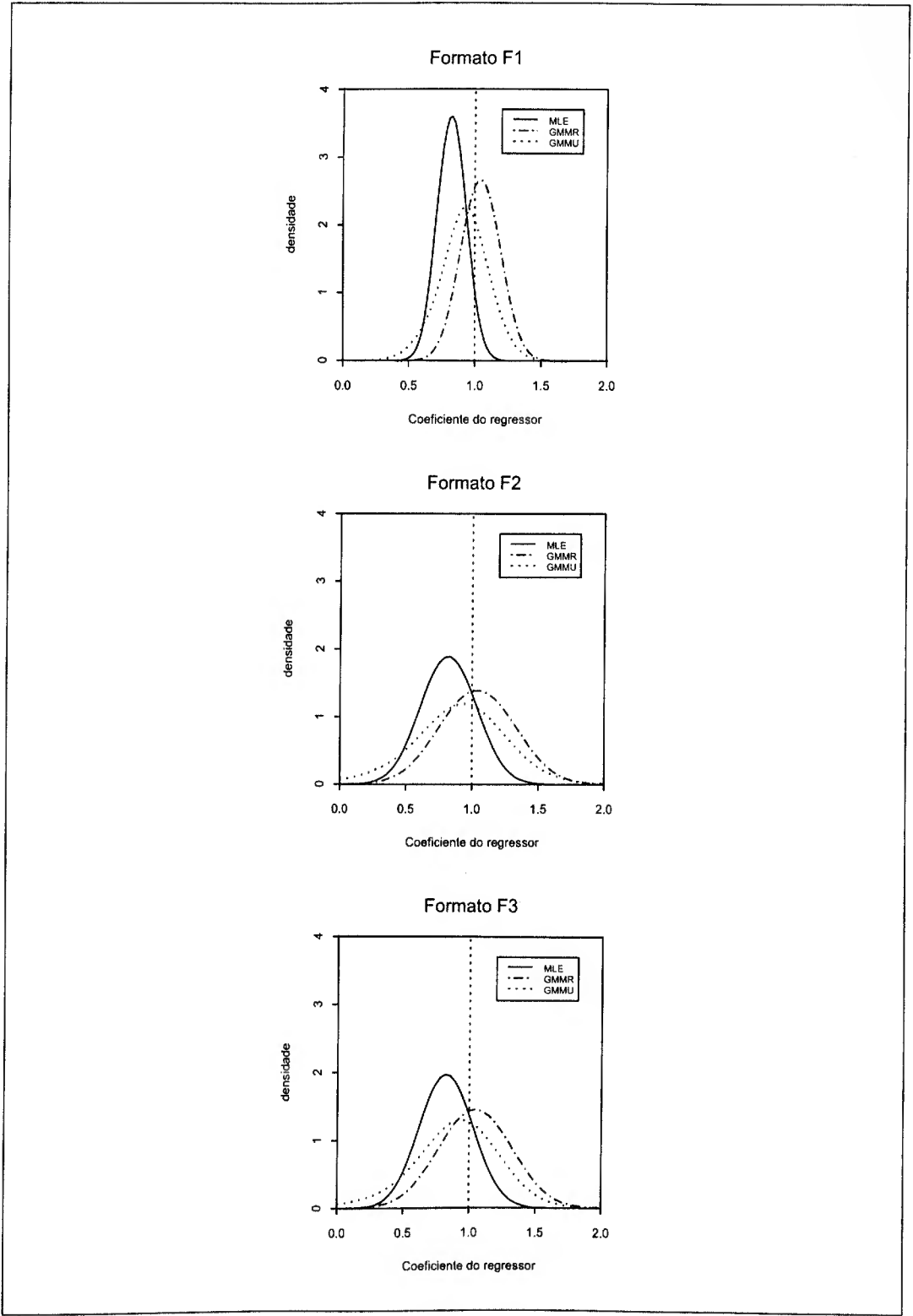
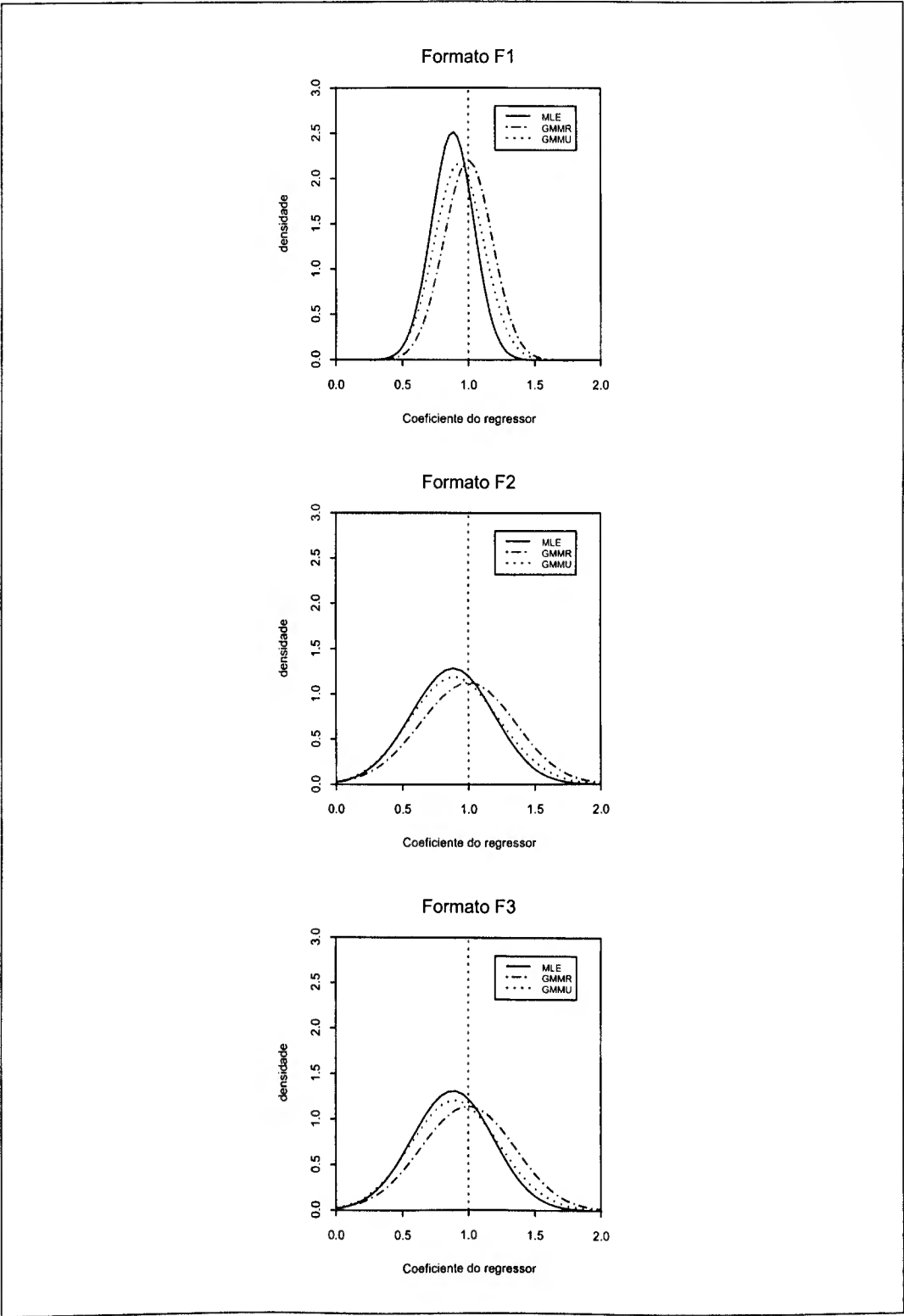


Figura 6.4: Densidades dos estimadores MLE, GMMR e GMMU da estimação de  $\beta_1$  no modelo Log-logistic com distribuição de heterogeneidade Lognormal e  $n = 500$



## 6.3 Modelo com observações censuradas

Vamos agora analisar as propriedades dos estimadores na presença de observações censuradas através de um mecanismo de censura aleatória. Para tal, vamos considerar uma distribuição como a da secção anterior, com  $\alpha = 1$ ,  $\{\beta_0, \beta_1\} = \{0, 1\}$  e  $\rho = 0.8$ . Para o conjunto de regressores  $x = [1, x_1]$  foi escolhido o formato F1.

Os dados foram gerados através de  $\{t_i, c_i, x_i, d_i\}_{i=1}^n$ , onde  $t_i$  representa a permanência do  $i$ -ésimo indivíduo no estado,  $c_i$  o tempo de censura e  $d_i$  o indicador que diz se a observação é censurada ou não.

Na presença de observações censuradas à direita, o valor da duração observada será dado por  $T_i = \min\{T_i, C_i\}$  sendo o indicador de censura obtido através de  $d_i = I(T_i < C_i)$ , tal como referido em (2.8).

Como o processo de censura é efectuado através de um processo aleatório, o tempo de censura  $C_i$  é uma variável aleatória independente da verdadeira duração  $T_i$ , com uma função sobrevivência dada por  $Q(c|\lambda) = S(c, \alpha, \omega\lambda)$ , onde  $\lambda = \exp(x'\beta)$  e  $w$  é uma constante escolhida de forma que a proporção de observações censuradas,  $\gamma$ , seja a pretendida. Assim,

$$\Pr(d = 0|C) = \Pr(T > C|\alpha, \lambda, \omega) = \gamma$$

Sendo  $Y = \frac{T}{C}$ , os valores de  $\omega$  são determinados através da resolução da equação

$$\Pr(Y > 1|\alpha, \lambda, \omega) = \gamma$$

É fácil verificar que  $Y$  é uma variável aleatória positiva com função densidade

$$f(y) = \frac{\alpha \omega y^{\alpha-1}}{(\omega + y^\alpha)^2}$$

para o modelo Weibull.e

$$f(y) = \frac{\alpha \omega y^{\alpha-1}(2\omega - 2y^\alpha + (\omega + y^\alpha)(\alpha \log y - \log \omega))}{(y^\alpha - \omega)^3}$$

para o modelo Log-logistic.

Tal como anteriormente, vamos ter de derivar as funções aproximadas para a condição de momentos corrigidos para os estimadores GMM. Considerando que  $g_{1\bullet}^c(t, x, \theta)$  corresponde às funções aproximadas para a condição de momentos corrigidos para observações completas dadas por (6.1) e (6.3) para o modelo Weibull e (6.2) e (6.4) para o modelo Log-logistic, essas funções, para dados censurados, para o modelo Weibull são dadas por

$$\begin{aligned} g_\alpha^c(t, x, \theta) &= d g_{1\alpha}^c(t, x, \theta) - \left( (1-d)z(c) + \frac{\sigma^2}{2} z(c) \exp(-z(c)) \right) \log(c) \\ g_\beta^c(t, x, \theta) &= x \left( d g_{1\beta}^c(t, x, \theta) - (1-d)z(c) - \frac{\sigma^2}{2} z(c) \exp(-z(c)) \right) \\ g_{\sigma^2}^c(t, x, \theta) &= d g_{1\sigma^2}^c(t, x, \theta) - (1-d)z(c)(1-z(c)) - \frac{\sigma^2}{2} z(c) \exp(-z(c)) \end{aligned} \quad (6.5)$$

e para o modelo Log-logistic por

$$\begin{aligned} g_\alpha^c(t, x, \theta) &= d g_{1\alpha}^c(t, x, \theta) - (1-d) \frac{z(c) \log(c)}{1+z(c)} - \frac{\sigma^2 z(c) \log(c)}{2(1+z(c))^3} \\ g_\beta^c(t, x, \theta) &= x \left( d g_{1\beta}^c(t, x, \theta) - (1-d) \frac{z(c)}{1+z(c)} - \frac{\sigma^2 z(c)}{2(1+z(c))^3} \right) \\ g_{\sigma^2}^c(t, x, \theta) &= d g_{1\sigma^2}^c(t, x, \theta) - (1-d) \frac{z(c)(1-z(c))}{(1+z(c))^2} - \frac{\sigma^2 z(c)(1+2z(c)-3z(c)^2)}{2(1+z(c))^5} \end{aligned} \quad (6.6)$$

A análise vai ser realizada para duas proporções de censura: 0.2 e 0.5.

Na tabela 6.10 são apresentados os resultado para o estimador de Máxima Verosimilhança:

Tabela 6.10: Resultados da simulação dos modelos Weibull e Log-logistic com censura para  $\rho=0.8$  e  $\alpha=1$ : estimador de Máxima Verosimilhança

			Máxima Verosimilhança							
			Weibull				Log-logistic			
			$\beta_1/\beta_1$		$\hat{\alpha}/\alpha$		$\beta_1/\beta_1$		$\hat{\alpha}/\alpha$	
Censura	n	U	média	dp	média	dp	média	dp	média	dp
0.20	200	LN	.832	.154	.833	.051	.878	.224	.873	.055
		G	.809	.167	.812	.058	.879	.225	.873	.056
	500	LN	.828	.098	.826	.033	.864	.143	.867	.035
		G	.806	.103	.803	.037	.867	.139	.868	.035
0.50	200	LN	.833	.191	.834	.061	.852	.254	.848	.067
		G	.807	.201	.811	.066	.849	.257	.843	.067
	500	LN	.828	.119	.826	.039	.841	.163	.842	.042
		G	.805	.124	.803	.042	.838	.160	.838	.042

- No modelo Log-logistic, podemos verificar que o enviesamento proporcional do *shape parameter* e do coeficiente do regressor vão aumentando à medida que aumenta a censura, ao contrário do modelo Weibull, onde a censura parece não ter influência no enviesamento dos parâmetros.

A tabela 6.11 apresenta os resultados da simulação para os modelos Weibull e Log-logistic quando se utiliza o estimador GMMR:

- Verifica-se, em ambos os modelos, uma melhoria significativa no enviesamento proporcional do *shape parameter* e no coeficiente do regressor.
- O enviesamento, quer do coeficiente do regressor, quer do *shape parameter*, é menor no modelo Log-logistic.
- O aumento da censura origina um aumento do enviesamento proporcional no coeficiente do regressor e no *shape parameter* em ambos os modelos, assim como uma diminuição na precisão.

Tabela 6.11: Resultados da simulação dos modelos Weibull e Log-logistic com censura para  $\rho=0.8$  e  $\alpha=1$ : estimador GMMR

			GMMR							
			Weibull				Log-logistic			
			$\beta_1/\beta_1$		$\hat{\alpha}/\alpha$		$\beta_1/\beta_1$		$\hat{\alpha}/\alpha$	
Censura	n	U	média	dp	média	dp	média	dp	média	dp
0.20	200	LN	1.021	.196	1.021	.069	1.011	.257	.994	.067
		G	.986	.216	.990	.079	1.012	.257	1.004	.064
	500	LN	1.013	.126	1.010	.044	.994	.164	.997	.040
		G	.979	.136	.975	.051	.998	.159	.999	.041
0.50	200	LN	.953	.224	.954	.076	.977	.289	.972	.077
		G	.919	.237	.925	.082	.973	.292	.965	.077
	500	LN	.945	.139	.943	.048	.963	.186	.964	.048
		G	.916	.146	.913	.052	.960	.182	.960	.048

A tabela 6.12 mostra-nos os resultados da simulação do modelo Weibull através do estimador GMM com  $\sigma^2$  desconhecido em presença de observações censuradas:

Tabela 6.12: Resultados da simulação o modelo Weibul com censura para  $\rho=0.8$  e  $\alpha=1$ : estimador GMMU

			GMMU					
			Weibull					
			$\beta_1/\beta_1$		$\hat{\alpha}/\alpha$		$\hat{\sigma}^2/\sigma^2$	
Censura	n	U	média	dp	média	dp	média	dp
0.20	200	LN	.869	.230	.926	.118	.582	.455
		G	.799	.278	.899	.125	.625	.441
	500	LN	.843	.169	.914	.090	.600	.330
		G	.730	.204	.856	.096	.521	.317
0.50	200	LN	.817	.244	.874	.116	.432	.543
		G	.744	.263	.834	.112	.378	.509
	500	LN	.774	.151	.839	.071	.279	.325
		G	.693	.172	.798	.067	.238	.308

- A censura tem uma forte influência no enviesamento proporcional do *shape parameter* e do coeficiente do regressor, aumentando o enviesamento à medida que aumenta a censura.
- No estimador da variância da heterogeneidade, verifica-se um aumento do seu enviesamento e uma diminuição da sua precisão, que é agravado com o aumento

da proporção de censura.

Na tabela 6.13 são apresentados os resultados para o modelo Log-logistic da simulação através do estimador GMM com  $\sigma^2$  desconhecido na presença de observações censuradas:

Tabela 6.13: Resultados da simulação do modelo Log-logistic com censura para  $\rho=0.8$  e  $\alpha=1$ : estimador GMMU

			GMMU					
			Log-logistic					
			$\beta_1/\beta_1$		$\hat{\alpha}/\alpha$		$\hat{\sigma}^2/\sigma^2$	
Censura	n	U	média	dp	média	dp	média	dp
0.20	200	LN	.940	.250	.942	.085	.526	.473
		G	.937	.248	.937	.087	.495	.477
	500	LN	.943	.169	.953	.065	.659	.386
		G	.952	.160	.958	.064	.687	.385
0.50	200	LN	.914	.280	.917	.092	.565	.477
		G	.904	.283	.909	.094	.551	.496
	500	LN	.921	.187	.930	.067	.733	.398
		G	.918	.184	.929	.066	.754	.376

- Tal como no modelo Weibull, com o aumento da censura, verifica-se um aumento do enviesamento em ambos os parâmetros. No entanto, os valores obtidos para os parâmetros no modelo Log-logistic têm um enviesamento menor que o verificado no modelo Weibull.
- O enviesamento da variância da heterogeneidade é muito superior ao verificado quando não existe censura.

Nas figuras 6.5 e 6.6 podemos ver a estimativa da densidade da estimação do *shape parameter* e do coeficiente do regressor em presença de observações censuradas, quando utilizamos, respectivamente, o modelo Weibull e o modelo Log-logistic.



Figura 6.5: Densidades dos estimadores MLE, GMMR e GMMU da estimação de  $\alpha$  e  $\beta_1$  no modelo Weibull com censura e com distribuição de heterogeneidade Lognormal e  $n = 500$

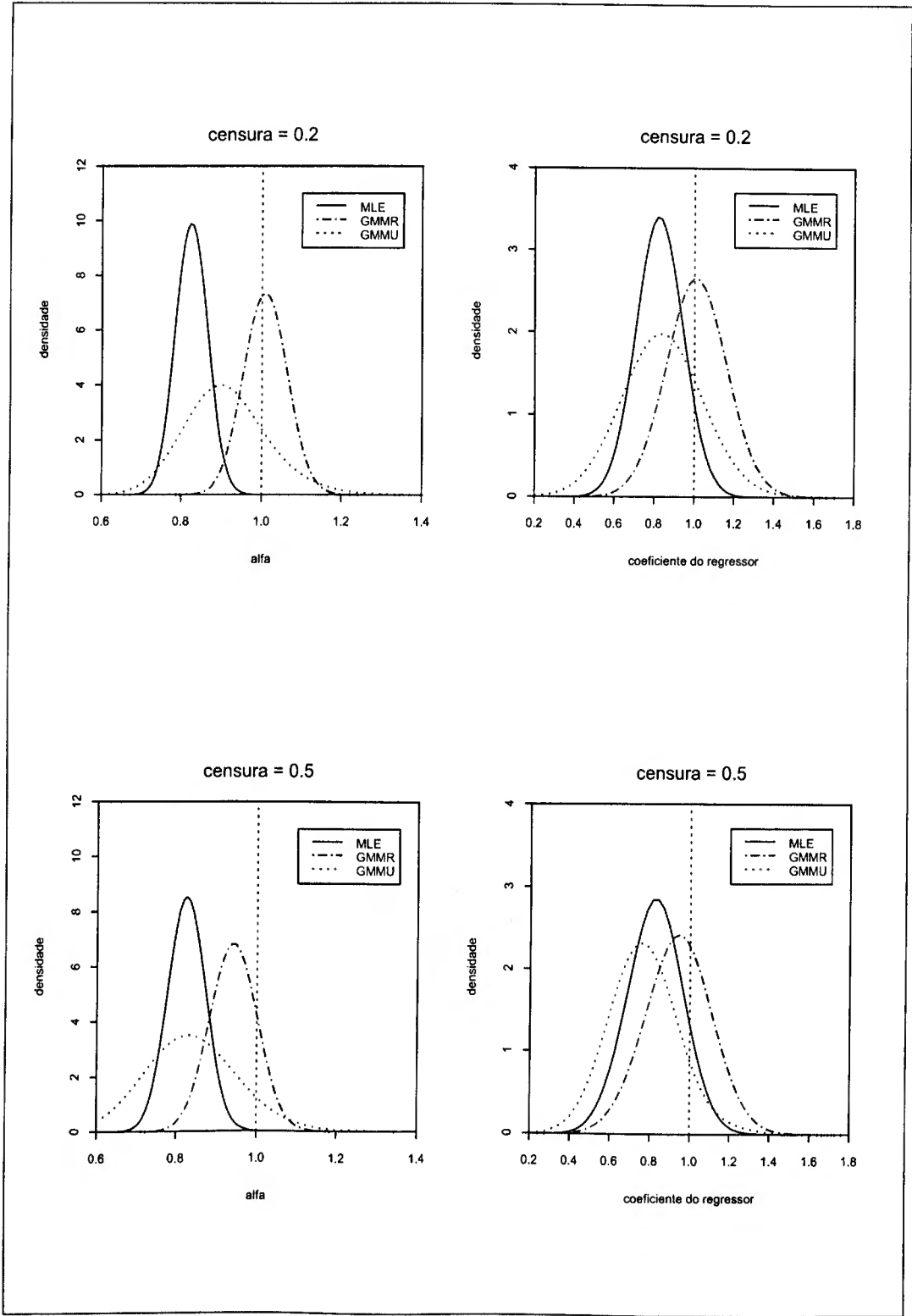
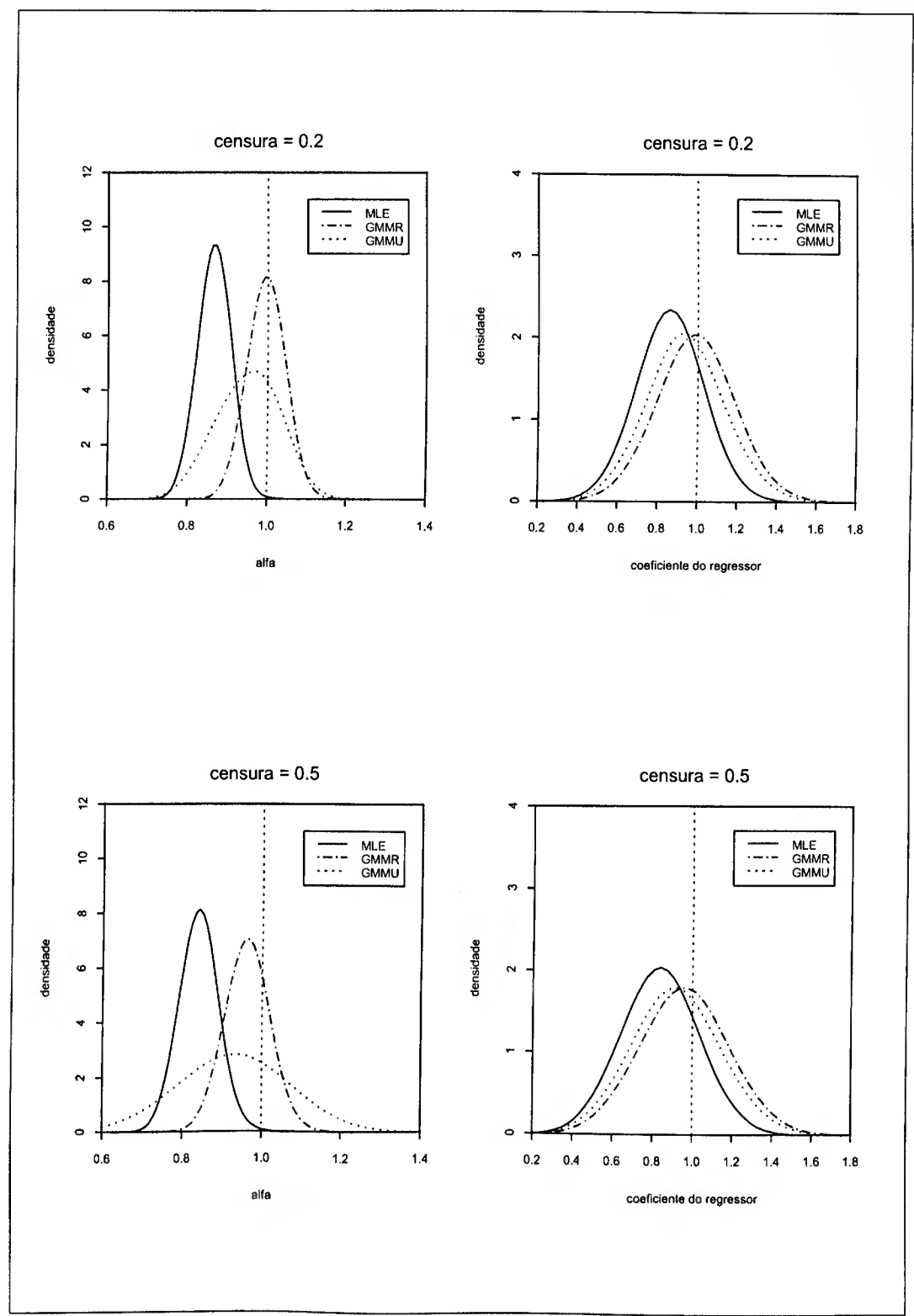


Figura 6.6: Densidades dos estimadores MLE, GMMR e GMMU da estimação de  $\alpha$  e  $\beta_1$  no modelo Log-logistic com censura e com distribuição de heterogeneidade Lognormal e  $n = 500$



## 7. Conclusões

Na análise econometrica de modelos de duração, é frequente encontrar-se o problema da heterogeneidade negligenciada. Este problema, que pode ser visto como uma omissão de variáveis explicativas, ou como erro de medida na variável dependente, origina enviesamento nos estimadores, quer na estimação dos parâmetros funcionais, quer nos coeficientes dos regressores.

Nos modelos de duração, ignorar a heterogeneidade na estimação, conduz quase sempre a estimativas enviesadas para os parâmetros do modelo. O estudo que aqui se conclui, tinha como principal objectivo analisar o efeito da heterogeneidade em modelos de duração com parâmetro de escala sem assumir uma forma funcional para a distribuição da heterogeneidade.

Considerando que a variância da heterogeneidade é pequena e utilizando aproximações assintóticas, foi possível construir uma função densidade semelhante à verdadeira densidade do modelo. Através da análise da aproximação da função *hazard*, podemos verificar que o efeito da heterogeneidade na função *hazard* não é constante ao longo da duração, estando dependente da forma funcional assumida para a função *hazard* e da própria duração. No entanto, para os modelos com parâmetro de escala pertencentes à família das funções *hazard* proporcionais, verificou-se que esse efeito era sempre na mesma direcção, originando que a função *hazard* com heterogeneidade tenha uma elasticidade inferior à função *hazard* onde esta é negligenciada.

Utilizando as aproximações assintóticas construímos uma estatística de teste para detecção da heterogeneidade negligenciada. O teste que foi proposto foi um teste score pertencente à classe dos testes de Matriz de Informação.

Com o objectivo de caracterizar o impacto da heterogeneidade na estimação

dos parâmetros do modelo, derivamos uma expressão para a aproximação do limite em probabilidade do estimador GMM que ignora a heterogeneidade definido por uma condição de momentos. A expressão obtida foi uma combinação linear do enviesamento na condição de momentos induzido pela heterogeneidade negligenciada.

Nos modelos propostos neste trabalho, Weibull e Log-logistic, verificou-se que o enviesamento proporcional do *shape parameter* é uma função das características funcionais do modelo de duração e independente dos coeficientes dos regressores.

Foi considerado um estimador GMM que corrige o enviesamento na condição de momentos induzido pela heterogeneidade. A função enviesamento é uma função estrutural visto incluir funções de  $T$  e de  $X$ , sendo derivada da função responsável pelo enviesamento no estimador GMM.

Através de simulações de Monte Carlo, estimamos os parâmetros para duas distribuições paramétricas, Weibull e Log-logistic, através da máxima verosimilhança e utilizando o estimador GMM corrigido. No modelo Log-logistic, verificou-se uma boa correcção do enviesamento dos parâmetros, sendo de referir que este modelo parece ser pouco sensível às diferentes distribuições da heterogeneidade aqui consideradas. Quanto ao modelo Weibull, a correcção efectuada pelo estimador GMM corrigido foi menor que a verificada para o modelo Log-logistic, sendo este modelo muito sensível à distribuição da heterogeneidade.

Com este trabalho, pretendeu-se mostrar alguma da utilidade das aproximações assintóticas na análise de modelos de duração com heterogeneidade negligenciada. Como demonstrado, para além de contribuírem para compreender o impacto da heterogeneidade neste tipo de modelos, as aproximações podem ser utilizadas para a sua detecção e, em alguns casos, contribuir para o desenvolvimento de métodos de estimação semiparamétrica.

## Apendice A

### Aproximação da função condicional de enviesamento - uma duração

Neste apêndice vamos derivar a aproximação da função enviesamento estrutural para modelos com observações completas e com observações censuradas.

Seja a aproximação da função densidade dada por

$$f_a(t) = f(t) + \frac{\sigma^2}{2\alpha^2} h(t) \quad (\text{A-1})$$

onde

$$h(t) = f(t) + 3t f'(t) + t^2 f''(t) \quad (\text{A-2})$$

e seja a aproximação da função condicional de enviesamento  $b(\phi)$  dada por

$$\int_0^\infty g(t, \phi) h(t, \phi) dt \quad (\text{A-3})$$

Vamos derivar a expressão para a aproximação da função estrutural de enviesamento de forma a que esta seja escrita em função da condição de momentos, ou seja, vamos determinar  $r(t, \phi)$  de forma que

$$\int_0^\infty g(t, \phi) h(t, \phi) dt = \int_0^\infty r(t, \phi) f(t, \phi) dt \quad (\text{A-4})$$

Substituindo (A-2) em (A-4), o lado esquerdo de (A-4) pode ser escrito como

$$\int_0^\infty g(t) f(t) dt + 3 \int_0^\infty t g(t) f'(t) dt + \int_0^\infty t^2 g(t) f''(t) dt$$

onde foi omitido o parâmetro  $\phi$  para simplificação da notação.

Resolvendo os dois últimos integrais através da integração por partes e simpli-

ficando, obtem-se a função

$$[3t g(t)f(t) + t^2 g(t)f'(t)]_0^\infty - 2 \int_0^\infty g(t)f(t)dt - 3 \int_0^\infty t g'(t)f(t)dt - \\ - 2 \int_0^\infty t g(t)f'(t)dt + \int_0^\infty t^2 g'(t)f'(t)dt$$

Fazendo uma nova integração por partes aos dois últimos integrais obtém-se à função

$$[t g(t)f(t) + t^2 g(t)f'(t) - t^2 g'(t)f(t)]_0^\infty + \int_0^\infty t g'(t)f(t)dt + \int_0^\infty t^2 g''(t)f(t)dt \quad (\text{A-5})$$

Assumindo que, devido à configuração das caudas das respectivas densidades, são válidas as seguintes condições:

$$\begin{aligned} \lim_{t \rightarrow 0} t g(t)f(t) &= 0 & \lim_{t \rightarrow \infty} t g(t)f(t) &= 0 \\ \lim_{t \rightarrow 0} t^2 g(t)f'(t) &= 0 & \lim_{t \rightarrow \infty} t^2 g(t)f'(t) &= 0 \\ \lim_{t \rightarrow 0} t^2 g'(t)f(t) &= 0 & \lim_{t \rightarrow \infty} t^2 g'(t)f(t) &= 0 \end{aligned} \quad (\text{A-6})$$

obtemos

$$b(\phi) = \frac{1}{2\alpha^2} \int_0^\infty (t g'(t, \phi) + t^2 g''(t, \phi))f(t, \phi)dt$$

sendo a aproximação da função estrutural de enviesamento dada por

$$b(t, \phi) = \frac{1}{2\alpha^2} [t g'(t, \phi) + t^2 g''(t, \phi)]$$

Para o caso em que existem observações censuradas, o procedimento será seme-

lhante: seja a aproximação da função sobrevivência dada por

$$S_a(t) = S(t) + \frac{\sigma^2}{2\alpha^2} H(t)$$

onde

$$H(t) = t S'(t) + t^2 S''(t) \quad (\text{A-7})$$

e seja a aproximação da função condicional de enviesamento  $b(\phi)$  dada por

$$\frac{1}{2\alpha^2} \left( \int_0^c g_1(t, \phi) h(t, \phi) dt + g_0(c, \phi) H(c, \phi) \right) \quad (\text{A-8})$$

Tal como no caso anterior, vamos derivar a expressão para a aproximação da função estrutural de enviesamento de forma a que seja escrita em função da condição de momentos, ou seja, vamos determinar  $r_1(t, \phi)$  de forma que

$$\int_0^c g_1(t, \phi) h(t, \phi) dt = \int_0^c r_1(t, \phi) f(t, \phi) dt + q(c, \phi) \quad (\text{A-9})$$

onde  $q(c, \phi) = q_1(c, \phi) + g_0(c, \phi) H(c, \phi)$

Substituindo (A-2) no lado esquerdo de (A-9) obtemos

$$\int_0^c g_1(t) f(t) dt + 3 \int_0^c t g(t) f'(t) dt + \int_0^c t^2 g_1(t) f''(t) dt$$

onde foi omitido o parâmetro  $\phi$  para simplificação da notação.

Fazendo a integração por partes, obtemos

$$[t g_1(t) f(t) + t^2 g_1(t) f'(t) - t^2 g_1'(t) f(t)]_0^c + \int_0^c t g_1'(t) f(t) dt + \int_0^c t^2 g_1''(t) f(t) dt$$

Assumindo novamente que devido à configuração das caudas das densidades, são válidas as condições definidas em (A-6), podemos escrever a expressão anterior

como

$$(c g_1(c) - c^2 g_1'(c))f(c) + c^2 g_1(c)f'(c) + \int_0^c t g_1'(t)f(t)dt + \int_0^c t^2 g_1''(t)f(t)dt$$

de forma que, tendo em conta que  $S'(c) = -f(c)$  e  $S''(c) = -f'(c)$  e considerando (A-7), as expressões de  $r_1(t, \phi)$  e  $q(c, \phi)$  são dadas por

$$r_1(t, \phi) = t g_1'(t, \phi) + t^2 g_1''(t, \phi)$$

$$q(c, \phi) = (c g_1(c) - c^2 g_1'(c) + c g_0(c))f(c) + (c^2 g_1(c) - c^2 g_0(c))f'(c)$$

sendo a aproximação da função condicional de enviesamento dada por

$$b(t, c, \phi) = \frac{1}{2\alpha^2}(r_1(t, \phi) + q(c, \phi))$$



## Apêndice B

### Aproximação da função condicional de enviesamento - múltiplas durações

Neste apêndice vamos derivar a aproximação da função enviesamento estrutural para modelos com múltiplas durações.

Tendo em conta a aproximação da função densidade (5.2) e substituindo em (5.5) e (5.6)<sup>8</sup>, obtemos as funções

$$\begin{aligned} b^k(\phi) = & \frac{1}{2\alpha_k^2} \left( \int_0^\infty \dots \int_0^\infty g(\mathbf{t}) f(\mathbf{t}) dt_R \dots dt_1 + \right. \\ & \left. + 3 \int_0^\infty \dots \int_0^\infty t_k g(\mathbf{t}) f^k(\mathbf{t}) dt_R \dots dt_1 + \int_0^\infty \dots \int_0^\infty t_k^2 g(\mathbf{t}) f^{kk}(\mathbf{t}) dt_R \dots dt_1 \right) \end{aligned} \quad (\text{B-1})$$

$$\begin{aligned} b^{kl}(\phi) = & \frac{1}{\alpha_k \alpha_l} \left( \int_0^\infty \dots \int_0^\infty g(\mathbf{t}) f(\mathbf{t}) dt_R \dots dt_1 + \int_0^\infty \dots \int_0^\infty t_k t_l g(\mathbf{t}) f^{kl}(\mathbf{t}) dt_R \dots dt_1 \right. \\ & \left. + \int_0^\infty \dots \int_0^\infty t_k g(\mathbf{t}) f^k(\mathbf{t}) dt_R \dots dt_1 + \int_0^\infty \dots \int_0^\infty t_l g(\mathbf{t}) f^l(\mathbf{t}) dt_R \dots dt_1 \right) \end{aligned} \quad (\text{B-2})$$

Assumindo que, devido à configuração das caudas das respectivas densidades, são válidas as seguintes condições:

$$\begin{array}{ll} \lim_{t \rightarrow 0} t_k g(\mathbf{t}) f(\mathbf{t}) = 0 & \lim_{t \rightarrow \infty} t_k g(\mathbf{t}) f(\mathbf{t}) = 0 \\ \lim_{t \rightarrow 0} t_k^2 g(\mathbf{t}) f^k(\mathbf{t}) = 0 & \lim_{t \rightarrow \infty} t_k^2 g(\mathbf{t}) f^k(\mathbf{t}) = 0 \\ \lim_{t \rightarrow 0} t_k^2 g^k(\mathbf{t}) f(\mathbf{t}) = 0 & \lim_{t \rightarrow \infty} t_k^2 g^k(\mathbf{t}) f(\mathbf{t}) = 0 \\ \lim_{t \rightarrow 0} t_k t_l g(\mathbf{t}) f^l(\mathbf{t}) = 0 & \lim_{t \rightarrow \infty} t_k t_l g(\mathbf{t}) f^l(\mathbf{t}) = 0 \\ \lim_{t \rightarrow 0} t_k t_l g^k(\mathbf{t}) f(\mathbf{t}) = 0 & \lim_{t \rightarrow \infty} t_k t_l g^k(\mathbf{t}) f(\mathbf{t}) = 0 \end{array}$$

---

<sup>8</sup> Para simplificação da notação, daqui em diante será omitido o parâmetro  $\phi$ .

e aplicando a integração múltipla por partes obtém-se as seguintes igualdades

$$\int_0^\infty \dots \int_0^\infty t_k g(\mathbf{t}) f^k(\mathbf{t}) dt_R \dots dt_1 = - \int_0^\infty \dots \int_0^\infty (g(\mathbf{t}) + t_k g^k(\mathbf{t})) f(\mathbf{t}) dt_R \dots dt_1$$

$$\begin{aligned} \int_0^\infty \dots \int_0^\infty t_{kk} g(\mathbf{t}) f^{kk}(\mathbf{t}) dt_R \dots dt_1 &= \int_0^\infty \dots \int_0^\infty (2g(\mathbf{t}) + 4t_k g^k(\mathbf{t}) + \\ &+ t_k^2 g^{kk}(\mathbf{t}) f(\mathbf{t})) dt_R \dots dt_1 \end{aligned}$$

$$\begin{aligned} \int_0^\infty \dots \int_0^\infty t_k t_l g(\mathbf{t}) f^{kl}(\mathbf{t}) dt_R \dots dt_1 &= \int_0^\infty \dots \int_0^\infty (g(\mathbf{t}) + t_k g^k(\mathbf{t}) + t_l g^l(\mathbf{t}) + \\ &+ t_k t_l g^{kl}(\mathbf{t}) f(\mathbf{t})) dt_R \dots dt_1 \end{aligned}$$

que substituindo em (B-1) e (B-2) obtém-se

$$\begin{aligned} b^k(\phi) &= \frac{1}{2\alpha_k^2} \left( \int_0^\infty \dots \int_0^\infty t_k g^k(\mathbf{t}) f(\mathbf{t}) dt_R \dots dt_1 + \int_0^\infty \dots \int_0^\infty t_k^2 g^{kk}(\mathbf{t}) f(\mathbf{t}) dt_R \dots dt_1 \right) \\ b^{kl}(\phi) &= \frac{1}{\alpha_k \alpha_l} \int_0^\infty \dots \int_0^\infty t_k t_l g^{kl}(\mathbf{t}) f(\mathbf{t}) dt_R \dots dt_1 \end{aligned}$$

A aproximação da função estrutural do enviesamento será dada por

$$b(\mathbf{t}, \phi) = \frac{1}{2} \sum_{k=1}^R \frac{\sigma_k^2}{\alpha_k^2} (t_k g^k(\mathbf{t}, \phi) + t_k^2 g^{kk}(\mathbf{t}, \phi)) + \sum_{k=1}^{R-1} \sum_{l=k+1}^R \frac{\sigma_{kl}}{\alpha_k \alpha_l} t_k t_l g^{kl}(\mathbf{t}, \phi)$$

## Bibliografia

- Abbring, J. H. e Van den Berg, G. J. (2003), *The unobserved heterogeneity distribution in duration analysis*, Working paper, Free University Amsterdam.
- Chesher, A. (1984), Testing for neglected heterogeneity, *Econometrica*, 52:4, pp. 865-872.
- Chesher, A. (1991), The effect of measurement error, *Biometrika*, 78:3, pp. 451-462.
- Chesher, A. (2000), *Improved GMM estimation under covariate measurement error*, Presented at the World Meeting of the Econometric Society, Seattle.
- Chesher, A., Dumangane, M. e Smith, R. J. (2002), Duration response measurement error, *Journal of econometrics*, 111, pp. 169-194.
- Chesher, A., Lancaster, T. e Irish (1985), On detecting the failure of distributional assumptions, *Annales de L'insee*, 59/60, pp. 7-45.
- Cox, D. R. (1972), Regression models and life tables (with discussion), *Journal of the Royal Statistical Society B*, 34, 187-220.
- Cox, D. R. (1975), Partial likelihood, *Biometrika*, 62:2, pp. 269-276.
- Devine, T. J e Kiefer, N. M. (1991), *Empirical labor economics: the search approach*, New York: Oxford University Press.
- Dumangane, M (2001), *Essays on duration response measurement error*, Phd Thesis, Department of Economics, Univerity of Bristol.
- Elbers, C. e Ridder, G. (1982), True and spurious duration dependence: The identifiability of the proportional hazard model, *Review of Economic Studies*, 49:3, pp. 403-409.
- Han, A. e Hausman, J. A. (1990), Flexible parametric estimation of duration and competing risk, *Journal of Applied Econometrics*, 5, pp. 1-28.
- Heckman, J. J. e Singer, B. (1984), Econometric duration analysis, *Journal of Econometrics*, 24, pp. 63-132.
- Heckman, J. J. e Singer, B. (1984a), The identifiability of the pproportional hazard model, *Review of Economic Studies*, 51, pp. 2311-241.
- Heckman, J. J. e Taber, C. R. (1994), Econometric mixture models and more general models for unobservables in duration analysis, *Statistical Methods in Medical Research*, 3, 279-302.

Horowitz, J. L. (1999), Semiparametric estimation of a proportional hazard model with unobserved heterogeneity, *Econometrica*, 67:5, pp. 1001-1028.

Hougaard, P. (1984), Life tables methods for heterogeneous populations: Distributions describing the heterogeneity, *Biometrika*, 71:1, pp. 75-83.

Hougaard, P. (1986), A class of multivariate failure time distributions, *Biometrika*, 73:3, pp. 671-678.

Hougaard, P. (2000), *Analysis of multivariate survival data*, New York: Springer.

Kalbfleisch, J. D. e Prentice, R. L. (1980), *The statistical analysis of failure time data*, New York: Wiley.

Kennan, J.F. (1985), The duration of contract strikes in U.S. manufacturing, *Journal of econometrics*, 28, pp. 5-28.

Kiefer, N. M. (1988), Economic duration data and hazard functions, *Journal of Economic Literature*, vol. XXVI, pp. 646-679.

Kiefer, N. M. e Neumann, G. R. (1989), *Search models and applied labor economics*, New York: Cambridge University Press.

Kiefer, N. M. e Skoog, G. R. (1984), Local asymptotic specification error analysis, *Econometrica*, 52:4, pp. 873-885.

Lancaster, T. (1979), Econometric methods for the duration of unemployment, *Econometrica*, 47:4, pp. 939-956.

Lancaster, T. (1985), Generalised residuals and heterogeneous duration models, *Journal of Econometrics*, 28, pp. 155-169.

Lancaster, T. (1990), *The econometric of transition data*. Cambridge, New York: Cambridge University Press.

Lillard, L. A. (1993), Simultaneous equations for hazards, *Journal of Econometrics*, 56, 189-217.

Lindstrom, D. P. (1996), Economic opportunity in Mexico and return migration from the United States, *Demography*, 33, pp. 357-374.

Meyer, B. D. (1995), *Semiparametric estimation of hazard models*, Working paper, Northwestern University, Evanston.

Murphy, S. A. (1994), Consistency in a proportional hazards model incorporating a random effect, *The Annals of Statistics*, 22:2, pp. 712-731.

Murphy, S. A. (1995) Asymptotic theory for the frailty model, *The Annals of*

*Statistics*, 23:1, pp. 192-198.

Narendranathan, W., Nickell, S. e Stern, J. (1985), Unemployment benefits revisited, *Economic Journal*, 95, pp. 307-329.

Neumann, G. R. (1997), Search models and duration data, in: M. H. Pesaran (ed.), *Handbook of Applied Econometrics: Microeconometrics*, Oxford: Basil Blackwell.

Van den Berg, G. J. (2001), Duration models: Specification, identification, and multiple destination, in: J. J. Heckman e E. Leamer (ed.), *Handbook of econometrics, Volume V*, Amsterdam: North-Holland.

Vaupel, Mantom and Stallard (1979), The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, 16, pp. 439-454.

White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica*, 50:1, pp. 1-25.

